PATENT ABSTRACTS OF JAPAN

(11)Publication number : **2002-091477**

(43)Date of publication of application : **27.03.2002**

(51)Int.Cl. **G10L 15/06**
G10L 15/14
G10L 15/18
G10L 15/28

(21)Application number : **2000-280674**

(22)Date of filing : **14.09.2000**

(71)Applicant : **MITSUBISHI ELECTRIC CORP**

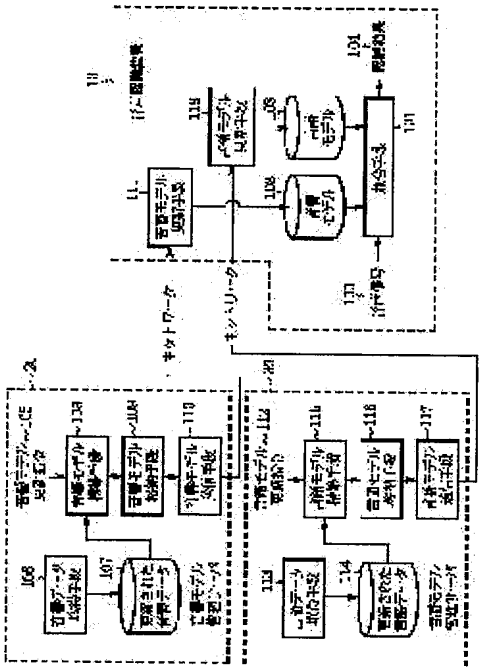
(72)Inventor : **OKATO YOHEI**
ISHII JUN

(54) VOICE RECOGNITION SYSTEM, VOICE RECOGNITION DEVICE, ACOUSTIC MODEL CONTROL SERVER, LANGUAGE MODEL CONTROL SERVER, VOICE RECOGNITION METHOD AND COMPUTER READABLE RECORDING MEDIUM WHICH RECORDS VOICE RECOGNITION PROGRAM

(57)Abstract:

PROBLEM TO BE SOLVED: To update acoustic and language models and to improve the precision in recognition without exerting a large load onto a user.

SOLUTION: An acoustic model control server 20 obtains updated acoustic data 107 and constructs an acoustic model. A language model control server 30 obtains updated language data 114 and constructs a language model. These models are respectively tansmitted to a voice recognition device 10. In the device 10, an acoustic model updating means 111 updates an acoustic model 102 by the transmitted acoustic model. A language model updating means 118 updates a language model 103 using the transmitted language model.



* NOTICES *

1.This document has been translated by computer. So the translation may not reflect the original precisely.
2.**** shows the word which can not be translated.
3.In the drawings, any words are not translated.

CLAIMS

[Claim(s)]
[Claim 1]Voice recognition equipment which inputs an audio signal, performs speech recognition with reference to an acoustic model which asks for probability of an audio acoustical observed value series, and outputs a recognition result.
An acoustic model managing server which is connected with the above-mentioned voice recognition equipment via a network, acquires updated sound data, and builds the above-mentioned acoustic model.
It updates by an acoustic model to which it is the voice recognition system provided with the above, the above-mentioned acoustic model which the above-mentioned acoustic model managing server built was transmitted to the above-mentioned voice recognition equipment, and the above-mentioned acoustic model managing server transmitted an acoustic model which the above-mentioned voice recognition equipment refers to in the case of speech recognition.

[Claim 2]It corresponds to a specific condition directed by ID from which an acoustic model managing server acquired and acquired ID which specifies an acoustic model which voice recognition equipment refers to in the case of speech recognition, The voice recognition system according to claim 1 reading updated sound data, building an acoustic model depending on the above-mentioned specific condition, and transmitting to the above-mentioned voice recognition equipment.
[Claim 3]Voice recognition equipment which inputs an audio signal, performs speech recognition with reference to a language model which searches for appearing probability of a word sequence, and outputs a recognition result.
A language model managing server which is connected with the above-mentioned voice recognition equipment via a network, acquires updated language data, and builds the above-mentioned language model.
It updates by a language model to which it is the voice recognition system provided with the above, the above-mentioned language model which the above-mentioned language model managing server built was transmitted to the above-mentioned voice recognition equipment, and the above-mentioned language model managing server transmitted a language model which the above-mentioned voice recognition equipment refers to in the case of speech recognition.

[Claim 4]It corresponds to a specific condition directed by ID from which a language model managing server acquired and acquired ID which specifies a language model which voice recognition equipment refers to in the case of speech recognition, The voice recognition system according to claim 3 reading updated language data, building a language model depending on the above-mentioned specific condition, and transmitting to the above-mentioned voice recognition equipment.
[Claim 5]Language data in which a language model managing server read the above-mentioned user dictionary, and was updated via a network with reference to a user dictionary with which voice recognition equipment registered a word on the occasion of speech recognition, The voice recognition system according to claim 3 building a language model depending on the above-mentioned user dictionary with reference to the read above-mentioned user dictionary, and transmitting to the above-mentioned voice recognition equipment.
[Claim 6]The voice recognition system according to claim 3, wherein a language model managing server acquires a text which a user of voice recognition equipment used, builds a language model depending on the above-mentioned text with reference to updated language data and the acquired above-mentioned text and transmits to the above-mentioned voice recognition equipment.
[Claim 7]Voice recognition equipment which inputs an audio signal, performs speech recognition with reference to an acoustic model which asks for probability of an audio acoustical observed value series, and outputs a recognition result.
An acoustic model managing server which is connected with the above-mentioned voice recognition equipment via a network, and has an initial acoustic model before adaptation.
ID as which it is the voice recognition system provided with the above, and the above-mentioned voice recognition equipment specifies the above-mentioned acoustic model, Acquire voice data for adaptation from an inputted audio signal, and voice data acquired ID and for adaptation, Transmit to the above-mentioned acoustic model managing server via a network, and the above-mentioned acoustic model managing server, Using transmitted voice data for adaptation, carry out adaptation of the above-mentioned initial acoustic model, and match and store an adaptation finishing acoustic model in above-mentioned ID to which it was transmitted, and. In response to renewal instructions of an acoustic model from the outside, ID which specifies the above-mentioned acoustic model from the above-mentioned voice recognition equipment via a network is received, An adaptation finishing acoustic model corresponding to ID which received is chosen and read out of a stored adaptation finishing acoustic model, It updates by an adaptation finishing acoustic model to which it transmitted to the above-mentioned voice recognition equipment via a network, and the above-mentioned acoustic model managing server transmitted an acoustic model which the above-mentioned voice recognition equipment refers to in the case of speech recognition.

[Claim 8]An acoustic model which asks for probability of an audio acoustical observed value series.
A collation means which inputs an audio signal, performs speech recognition with reference to the above-mentioned acoustic model, and outputs a recognition result.
From an acoustic model managing server which is voice recognition equipment provided with the above, and was connected via a network. It had an acoustic model update means updated by an acoustic model which received an acoustic model built with updated sound data, and received an acoustic model which the above-mentioned collation means refers to in the case of speech recognition.

[Claim 9]An acoustic model update means from an acoustic model managing server connected via a network. The voice recognition equipment according to claim 8 updating by an acoustic model which received an acoustic model depending on a specific condition of an acoustic model which a collation means refers to in the case of speech recognition built with updated sound data, and received an acoustic model which the above-mentioned collation means refers to in the case of speech recognition.
[Claim 10]A language model which searches for appearing probability of a word sequence.
A collation means which inputs an audio signal, performs speech recognition with reference to the above-mentioned language model, and outputs a recognition result.
From a language model managing server which is voice recognition equipment provided with the above, and was connected via a network. It had a

language model update means updated by a language model which received a language model built with updated language data, and received a language model which the above-mentioned collation means refers to in the case of speech recognition.

[Claim 11]A language model update means from a language model managing server connected via a network. The voice recognition equipment according to claim 10 updating by a language model which received a language model depending on a specific condition of a language model which a collation means refers to in the case of speech recognition built with updated language data, and received a language model which the above-mentioned collation means refers to in the case of speech recognition.
[Claim 12]A collation means is provided with a user dictionary which registered a word referred to in the case of speech recognition, A language model update means from a language model managing server connected via a network. The voice recognition equipment according to claim 10 updating by a language model which received a language model depending on a user dictionary which was built with updated language data, and which the above-mentioned collation means refers to in the case of speech recognition, and received a language model which the above-mentioned collation means refers to in the case of speech recognition.
[Claim 13]A language model update means from a language model managing server connected via a network. The voice recognition equipment according to claim 10 updating by a language model which received a language model depending on a text which was built with updated language data, and which a user who performs speech recognition used, and received a language model which the above-mentioned collation means refers to in the case of speech recognition.
[Claim 14]Voice recognition equipment comprising:
An acoustic model which asks for probability of an audio acoustical observed value series.
A collation means which inputs an audio signal, performs speech recognition with reference to the above-mentioned acoustic model, and outputs a recognition result.
An acoustic model ID acquiring means which acquires ID which specifies the above-mentioned acoustic model.
Read ID which the above-mentioned acoustic model ID acquiring means acquired, and voice data for adaptation is acquired from an inputted audio signal, A voice acquisition means for adaptation which transmits voice data read ID and for acquired adaptation to an acoustic model managing server to which it was connected via a network, An acoustic model update means updated by an adaptation finishing acoustic model which received an adaptation finishing acoustic model by which adaptation was carried out with voice data for the above-mentioned adaptation corresponding to above-mentioned ID from the above-mentioned acoustic model managing server, and received an acoustic model which the above-mentioned collation means refers to in the case of speech recognition.

[Claim 15]An acoustic model managing server comprising:
A sound data acquisition means which acquires updated sound data.
An acoustic model construction means which builds an acoustic model which reads updated sound data which the above-mentioned sound data acquisition means acquired in response to renewal instructions of an acoustic model from the outside, and asks for probability of an audio acoustical observed value series.
An acoustic model transmitting means which transmits an acoustic model built by the above-mentioned acoustic model construction means to voice recognition equipment which performs speech recognition via a network.

[Claim 16]An acoustic model managing server comprising:
A sound data acquisition means which acquires updated sound data.
An updating acoustic model ID acquiring means which acquires ID which specifies an acoustic model which voice recognition equipment connected via a network refers to in response to renewal instructions of an acoustic model from the outside in the case of speech recognition.
A sound data reading means for the specification which reads updated sound data which the above-mentioned sound data acquisition means acquired corresponding to a specific condition directed by ID which the above-mentioned updating acoustic model ID acquiring means acquired.
An acoustic model construction means for the specification which builds an acoustic model depending on the above-mentioned specific condition with reference to updated sound data which the above-mentioned sound data reading means for specification read, An acoustic model transmitting means which transmits an acoustic model which the above-mentioned acoustic model construction means for specification built to the above-mentioned voice recognition equipment via a network.

[Claim 17]An acoustic model managing server comprising:
An initial acoustic model before adaptation which asks for probability of an audio acoustical observed value series.
Voice data for adaptation transmitted from voice recognition equipment connected via a network.
An acoustic model adaptation means for the above-mentioned voice recognition equipment to receive ID which specifies an acoustic model referred to in the case of speech recognition, to carry out adaptation of the above-mentioned initial acoustic model using voice data for the above-mentioned adaptation, to match an adaptation finishing acoustic model with above-mentioned ID which received, and to store in an adaptation finishing acoustic model storing means.
In response to renewal instructions of an acoustic model from the outside, above-mentioned ID is received from the above-mentioned voice recognition equipment via a network, An adaptation finishing acoustic model selecting means which chooses and reads an adaptation finishing acoustic model corresponding to ID which received from the above-mentioned adaptation finishing acoustic model storing means, An acoustic model transmitting means which transmits an adaptation finishing acoustic model which the above-mentioned adaptation finishing acoustic model selecting means read to the above-mentioned voice recognition equipment via a network.

[Claim 18]A language model managing server comprising:
A language data acquisition means which acquires updated language data.
A language model construction means which builds a language model which reads updated language data which the above-mentioned language data acquisition means acquired in response to renewal instructions of a language model from the outside, and searches for appearing probability of a word sequence.
A language model transmitting means which transmits a language model which the above-mentioned language model construction means built to voice recognition equipment which performs speech recognition via a network.

[Claim 19]A language model managing server comprising:
A language data acquisition means which acquires updated language data.
An updating language model ID acquiring means which acquires ID which specifies a language model which voice recognition equipment connected via a network refers to in response to renewal instructions of a language model from the outside in the case of speech recognition.
A language data reading means for the specification which reads updated language data which the above-mentioned language data acquisition means acquired corresponding to a specific condition directed by ID which the above-mentioned updating language model ID acquiring means acquired.
A language model construction means for the specification which builds a language model depending on the above-mentioned specific condition with reference to updated language data which the above-mentioned language data reading means for specification read, A language model transmitting

means which transmits a language model which the above-mentioned language model construction means for specification built to the above-mentioned voice recognition equipment via a network.

[Claim 20]A language model managing server comprising:
A language data acquisition means which acquires updated language data.
A user dictionary reading means which reads a user dictionary which voice recognition equipment connected via a network refers to in response to renewal instructions of a language model from the outside in the case of speech recognition.
A user dictionary dependence language model construction means which builds a language model depending on a user dictionary which read updated language data which the above-mentioned language data acquisition means acquired, and the above-mentioned user dictionary reading means read.
A language model transmitting means which transmits a language model which the above-mentioned user dictionary dependence language model construction means built to the above-mentioned voice recognition equipment via a network.

[Claim 21]A language model managing server comprising:
A language data acquisition means which acquires updated language data.
A user use text acquisition means which acquires a text which a user of voice recognition equipment connected via a network used in response to renewal instructions of a language model from the outside.
A user use text dependence language model construction means which builds a language model depending on a text which read updated language data which the above-mentioned language data acquisition means acquired, and the above-mentioned user use text acquisition means acquired.
A language model transmitting means which transmits a language model which the above-mentioned user use text dependence language model construction means built to the above-mentioned voice recognition equipment via a network.

[Claim 22]A voice recognition method which inputs an audio signal, performs speech recognition with reference to an acoustic model which asks for probability of an audio acoustical observed value series, and outputs a recognition result, comprising:
The 1st step that acquires updated sound data.
The 2nd step that reads updated sound data which was acquired at the 1st step of the above in response to renewal instructions of an acoustic model, and builds an acoustic model.
The 3rd step that transmits an acoustic model built at the 2nd step of the above via a network.
The 4th step updated by an acoustic model which received an acoustic model which transmitted at the 3rd step of the above, and received an acoustic model referred to in the case of the above-mentioned speech recognition.

[Claim 23]A voice recognition method which inputs an audio signal, performs speech recognition with reference to a language model which searches for appearing probability of a word sequence, and outputs a recognition result, comprising:
The 1st step that acquires updated language data.
The 2nd step that reads updated language data which was acquired at the 1st step of the above in response to renewal instructions of a language model, and builds a language model.
The 3rd step that transmits a language model built at the 2nd step of the above via a network.
The 4th step updated by a language model which received a language model which transmitted at the 3rd step of the above, and received a language model referred to in the case of the above-mentioned speech recognition.

[Claim 24]A voice recognition method which inputs an audio signal, performs speech recognition with reference to an acoustic model which asks for probability of an audio acoustical observed value series, and outputs a recognition result, comprising:
The 1st step that acquires updated sound data.
The 2nd step that acquires ID which specifies an acoustic model referred to in the case of speech recognition in response to renewal instructions of an acoustic model.
The 3rd step that reads updated sound data which was acquired at the 1st step of the above corresponding to a specific condition directed by ID acquired at the 2nd step of the above.
The 4th step that builds an acoustic model depending on the above-mentioned specific condition with reference to updated sound data which was read at the 3rd step of the above.
The 5th step that transmits an acoustic model built at the 4th step of the above via a network.
The 6th step updated by an acoustic model which received an acoustic model which transmitted at the 5th step of the above, and received an acoustic model referred to in the case of speech recognition.

[Claim 25]A voice recognition method which inputs an audio signal, performs speech recognition with reference to a language model which searches for appearing probability of a word sequence, and outputs a recognition result, comprising:
The 1st step that acquires updated language data.
The 2nd step that acquires ID which specifies a language model referred to in the case of speech recognition in response to renewal instructions of a language model.
The 3rd step that reads updated language data which was acquired at the 1st step of the above corresponding to a specific condition directed by ID acquired at the 2nd step of the above.
The 4th step that builds a language model depending on the above-mentioned specific condition with reference to updated language data which was read at the 3rd step of the above.
The 5th step that transmits a language model built at the 4th step of the above via a network.
The 6th step updated by a language model which received a language model which transmitted at the 5th step of the above, and received a language model referred to in the case of speech recognition.

[Claim 26]A voice recognition method which inputs an audio signal characterized by comprising the following, performs speech recognition with reference to a language model which searches for appearing probability of a word sequence, and a user dictionary which registered a word, and outputs a recognition result.
The 1st step that acquires updated language data.
The 2nd step that reads a user dictionary referred to in the case of speech recognition in response to renewal instructions of a language model.
The 3rd step that builds a language model depending on a user dictionary which read updated language data which was acquired at the 1st step of the above, and was read at the 2nd step of the above.
The 4th step that transmits a language model built at the 3rd step of the above via a network.
The 5th step updated by a language model which received a language model which transmitted at the 4th step of the above, and received a language model referred to in the case of speech recognition.

[Claim 27]A voice recognition method which inputs an audio signal, performs speech recognition with reference to a language model which searches for appearing probability of a word sequence, and outputs a recognition result, comprising:
The 1st step that acquires updated language data.
The 2nd step that acquires a text which a user who performs speech recognition used in response to renewal instructions of a language model.
The 3rd step that builds a language model depending on a text which read updated language data which was acquired at the 1st step of the above, and was acquired at the 2nd step of the above.
The 4th step that transmits a language model built at the 3rd step of the above via a network.
The 5th step updated by a language model which received a language model which transmitted at the 4th step of the above, and received a language model referred to in the case of speech recognition.

[Claim 28]A voice recognition method which inputs an audio signal, performs speech recognition with reference to an acoustic model which asks for probability of an audio acoustical observed value series, and outputs a recognition result, comprising:
The 1st step that acquires ID which specifies the above-mentioned acoustic model.
The 2nd step that transmits voice data ID which read ID acquired at the 1st step of the above, acquired voice data for adaptation from an inputted audio signal, and was read via a network, and for acquired adaptation.
The 3rd step that carries out adaptation of the initial acoustic model before adaptation, and matches and stores an adaptation finishing acoustic model in ID which transmitted at the 2nd step of the above using voice data for adaptation transmitted at the 2nd step of the above.
The 4th step that receives ID acquired at the 1st step of the above via a network, and chooses and reads an adaptation finishing acoustic model corresponding to ID which received in response to renewal instructions of an acoustic model out of an adaptation finishing acoustic model stored at the 3rd step of the above.
The 5th step that transmits an adaptation finishing acoustic model read at the 4th step of the above via a network.
The 6th step updated by an adaptation finishing acoustic model which received an adaptation finishing acoustic model which transmitted at the 5th step of the above, and received an acoustic model referred to in the case of speech recognition.

[Claim 29]Input an audio signal and speech recognition is performed with reference to an acoustic model which asks for probability of an audio acoustical observed value series, A sound data acquisition function which is the recording medium which recorded a voice recognition program which realizes a verification function which outputs a recognition result, and acquires updated sound data, An acoustic model construction function to read updated sound data which the above-mentioned sound data acquisition function acquired in response to renewal instructions of an acoustic model, and to build an acoustic model, An acoustic model transmitting function which transmits an acoustic model which the above-mentioned acoustic model construction function built via a network, A recording medium which recorded a voice recognition program which realizes an acoustic model update function updated by an acoustic model which received an acoustic model which the above-mentioned acoustic model transmitting function transmitted, and received an acoustic model which the above-mentioned verification function refers to in the case of speech recognition and in which computer reading is possible.
[Claim 30]Input an audio signal and speech recognition is performed with reference to a language model which searches for appearing probability of a word sequence, A language data acquisition function which is the recording medium which recorded a voice recognition program which realizes a verification function which outputs a recognition result, and acquires updated language data, A language model construction function to read updated language data which the above-mentioned language data acquisition function acquired in response to renewal instructions of a language model, and to build a language model, A language model transmitting function which transmits a language model which the above-mentioned language model construction function built via a network, A recording medium which recorded a voice recognition program which realizes a language model update function updated by a language model which received a language model which the above-mentioned language model transmitting function transmitted, and received a language model which the above-mentioned verification function refers to in the case of speech recognition and in which computer reading is possible.
[Claim 31]Input an audio signal and speech recognition is performed with reference to an acoustic model which asks for probability of an audio acoustical observed value series, A sound data acquisition function which is the recording medium which recorded a voice recognition program which realizes a verification function which outputs a recognition result, and acquires updated sound data, An updating acoustic model ID acquisition function which acquires ID which specifies the above-mentioned acoustic model in response to renewal instructions of an acoustic model, It corresponds to a specific condition directed by ID which the above-mentioned updating acoustic model ID acquisition function acquired, A sound data read-out function for the specification which reads updated sound data which the above-mentioned sound data acquisition function acquired, An acoustic model construction function for the specification which builds an acoustic model depending on the above-mentioned specific condition with reference to updated sound data which the above-mentioned sound data read-out function for specification read, An acoustic model transmitting function which transmits an acoustic model which the above-mentioned acoustic model construction function for specification built via a network, A recording medium which recorded a voice recognition program which realizes an acoustic model update function updated by an acoustic model which received an acoustic model which the above-mentioned acoustic model transmitting function transmitted, and received an acoustic model which the above-mentioned verification function refers to in the case of speech recognition and in which computer reading is possible.
[Claim 32]Input an audio signal and speech recognition is performed with reference to a language model which searches for appearing probability of a word sequence, A language data acquisition function which is the recording medium which recorded a voice recognition program which realizes a verification function which outputs a recognition result, and acquires updated language data, An updating language model ID acquisition function which acquires ID which specifies the above-mentioned language model in response to renewal instructions of a language model, It corresponds to a specific condition directed by ID which the above-mentioned updating language model ID acquisition function acquired, A language data read-out function for the specification which reads updated language data which the above-mentioned language data acquisition function acquired, A language model construction function for the specification which builds a language model depending on the above-mentioned specific condition with reference to updated language data which the above-mentioned language data read-out function for specification read, A language model transmitting function which transmits a language model which the above-mentioned language model construction function for specification built via a network, A recording medium which recorded a voice recognition program which realizes a language model update function updated by a language model which received a language model which the above-mentioned language model transmitting function transmitted, and received a language model which the above-mentioned verification function refers to in the case of speech recognition and in which computer reading is possible.
[Claim 33]Input an audio signal and speech recognition is performed with reference to a language model which searches for appearing probability of a word sequence, and a user dictionary which registered a word, A language data acquisition function which is the recording medium which recorded a voice recognition program which realizes a verification function which outputs a recognition result, and acquires updated language data, A user dictionary read-out function which reads the above-mentioned user dictionary in response to renewal instructions of a language model, A user dictionary dependence language model construction function to build a language model depending on a user dictionary which read updated language data which the above-mentioned language data acquisition function acquired, and the above-mentioned user dictionary read-out function read, A language model transmitting function which transmits a language model which the above-mentioned user dictionary dependence language model construction function built via a network, A recording medium which recorded a voice recognition program which realizes a language model update function updated by a language model which received a language model which the above-mentioned language model transmitting function transmitted, and received a language model which the above-mentioned verification function refers to in the case of speech recognition and in which computer

reading is possible.
[Claim 34]Input an audio signal and speech recognition is performed with reference to a language model which searches for appearing probability of a word sequence, A language data acquisition function which is the recording medium which recorded a voice recognition program which realizes a verification function which outputs a recognition result, and acquires updated language data, A user use text acquisition function which acquires a text which a user who performs speech recognition used in response to renewal instructions of a language model, A user use text dependence language model construction function to build a language model depending on a text which read updated language data which the above-mentioned language data acquisition function acquired, and the above-mentioned user use text acquisition function acquired, A language model transmitting function which transmits a language model which the above-mentioned user use text dependence language model construction function built via a network, A recording medium which recorded a voice recognition program which realizes a language model update function updated by a language model which received a language model which the above-mentioned language model transmitting function transmitted, and received a language model which the above-mentioned verification function refers to in the case of speech recognition and in which computer reading is possible.
[Claim 35]Input an audio signal and speech recognition is performed with reference to an acoustic model which asks for probability of an audio acoustical observed value series, An acoustic model ID acquisition function which is the recording medium which recorded a voice recognition program which realizes a verification function which outputs a recognition result, and acquires ID which specifies the above-mentioned acoustic model, Read ID which the above-mentioned acoustic model ID acquisition function acquired, acquire voice data for adaptation from an inputted audio signal, and via a network, A voice acquisition function for adaptation which transmits voice data read ID and for acquired adaptation, Voice data for adaptation which the above-mentioned voice acquisition function for adaptation transmitted is used, An acoustic model adaptation function to carry out adaptation of the initial acoustic model before adaptation, and to match and store an adaptation finishing acoustic model in ID which the above-mentioned voice acquisition function for adaptation transmitted, In response to renewal instructions of an acoustic model, ID which the above-mentioned acoustic model ID acquisition function acquired via a network is received, An adaptation finishing acoustic model which an adaptation finishing acoustic model function preselection capability which chooses and reads an adaptation finishing acoustic model corresponding to ID which received out of an adaptation finishing acoustic model which the above-mentioned acoustic model adaptation function stored, and the above-mentioned adaptation finishing acoustic model function preselection capability read, An adaptation finishing acoustic model which an acoustic model transmitting function which transmits via a network, and the above-mentioned acoustic model transmitting function transmitted is received, A recording medium which recorded a voice recognition program which realizes an acoustic model update function in which the above-mentioned verification function updates an acoustic model referred to in the case of speech recognition by an adaptation finishing acoustic model which received and in which computer reading is possible.

[Translation done.]

DETAILED DESCRIPTION

[Detailed Description of the Invention]
[0001]
[Field of the Invention]This invention the acoustic model and language model which are referred to in the case of speech recognition so that a high recognition rate may be acquired, It is related with the recording medium which recorded the voice recognition system, the voice recognition equipment, the acoustic model managing server, the language model managing server, voice recognition method, and voice recognition program which are updated in the newest state via a network and in which computer reading is possible.
[0002]
[Description of the Prior Art]In speech recognition, after usually changing the sound digitized and inputted into the time series of the vector which expresses the audio acoustical feature well using the signal-processing technique, collation processing with an audio model (an acoustic model, a language model) is performed.
[0003]Collation processing is a problem which asks for word sequence W= [w1w2::wk] (k is the number of words) uttered from sound feature vector time series A= [a1a2::an] which consists of n time frames. In order to presume a word sequence as for which recognition precision becomes the highest, the appearing probability P (W|A) should just ask for recognized word sequence W* used as the maximum. Namely, W*=argmaxWP(W|A) (1)
However, it is usually difficult to ask for P (W|A) directly. Then, P (W|A) is rewritten like (2) types using the Bayes theorem.
P(W|A)=P(W)P(A|W)/P(A) (2)
[0004]Here, when asking for W which maximizes the left side, since P (A) which is a right-hand side denominator does not affect W which becomes a recognition candidate, it should just ask for W which maximizes a right-hand side numerator. That is, it becomes like (3) types.
W*=argmaxWP(W)P(A|W)(3)
Here, the stochastic model which gives P (W) is called a language model, and the stochastic model which gives P (A|W) is called an acoustic model.
[0005]These typical modeling methods in speech recognition are the methods of expressing an acoustic model by Hidden Markov Model, and expressing a language model by n-1-fold Markov process of the word called n gram.
[0006]The details of these methods, for example "Basic [of speech recognition] (top, under)" L.RABINER, B.H.JUANG, the Furui supervision of translation, November, 1995, NTT Advanced Technology (it is hereafter considered as the literature 1), It is describing in Shokodo (it is hereafter considered as the literature 3) in the "stochastic language model" Kitasato Institute 2, University of Tokyo Press (it is hereafter considered as the literature 2), "digital signal processing of sound and sound information" Kiyohiro Kano, Tetsu Nakamura, Shiro Ise collaboration, and November, 1997.
[0007]In these methods, the parameter which constitutes a model is statistically presumed from a lot of data. In construction of an acoustic model, voice data, such as a word from many speakers and a sentence, is collected beforehand, and it is presumed that the index well connected with recognition precision or recognition precision using the statistical method improves. For example, the parameter of the Hidden Markov Model which constitutes an acoustic model is presumed using a BAUMU Welch algorithm so that the likelihood to which an acoustic model outputs learned data may serve as the maximum. The estimation method of the acoustic model is described in detail in literature 1 beaming.
[0008]Similarly, in construction of a language model, the probability that the word which constitutes each utterance and utterance will appear from texts, such as a newspaper and the opening sentence of conversation, according to the structure of a language model is calculated. For example, in n gram language model, when it sets with n= 2 (called a bigram language model), P (W) is approximated like (4) types. The parameter of n gram language model is presumed from the frequency of adjoining n word in the text data for study. The estimation method of the language model is described in detail in the literature 2.
P(w1..w.k)
=P(w1)P(w2|w1)..P.(wk|w1..w.k-1)
**P(w1) P (w2|w1) .. P. (wk|wk-1) (4)
[0009]Thus, the language model which acquires high recognition precision compared with the method of not using a statistics value can be built by presuming the appearing probability of each word statistically using an extensive text. Since a text is not written with a space between words in Japanese, the definition of a word is ambiguous, but in the text, each unit which divided the text by a certain compatible means is defined as a word. This word is the division of a text based on linguistic units, such as a character, a morpheme, a clause, or an entropy standard, such combination, etc., for example.
[0010]Drawing 12 is a block diagram showing the composition of the conventional voice recognition equipment indicated by the above-mentioned literature 1. In drawing 12, the acoustic model referred to when the collation means 101 carries out speech recognition of the collation means which 101 inputs and carries out speech recognition of the audio signal 100, and outputs the recognition result 104, and 102, and 103 are language models referred to when the collation means 101 carries out speech recognition.
[0011]Next, operation is explained. The collation means 101 inputs a user's audio signal 100, performs speech recognition with reference to the acoustic model 102 and the language model 103, and outputs the recognition result 104. The acoustic model 102 expresses the mapping relation of the time series of the sound feature vector produced by carrying out signal processing of the voice waveform of a user's inputted audio signal 100, and the minimum symbol information that the voice recognition equipment expressed with a phoneme etc. treats, for example. The language model 103 describes the appearance information of a word to be a correspondence relation with recognition units longer than the word etc. which are expressed with the combination of the symbol mapped by the acoustic model 102. The acoustic model 102 asks for the probability that the model of a certain symbol will output a vector time series.
That is, it asks for the probability of an audio acoustical observed value series, and the language model 103 searches for the appearing probability of a certain word sequence.

[0012]Drawing 13 is a flow chart which shows the procedure of the conventional speech recognition processing in which input the audio signal 100 and the recognition result 104 is obtained. In step ST1301, the A/D conversion of the inputted audio signal 100 is carried out, and it turns into a digital signal. In step ST1302, the digitized audio signal sets a suitable interval, and signal processing is carried out, and it is changed into the time series of the sound feature vector which expresses audio character well.
[0013]In step ST1303, a sound feature vector is compared with the acoustic model 102 by sound collation processing, and the probability which outputs the time series of a sound feature vector about each recognition candidate is called for. In step ST1304, further, each recognition candidate is compared with the language model 103 by language collation processing, and the output probability of a word sequence can multiply by him by it. Finally,

in step ST1305, the most suitable candidate is chosen from each recognition candidate, and the recognition result 104 is obtained. Usually, the most suitable recognition result is the recognition candidate made the highest [probability] by the above-mentioned collation.
[0014]Only by the acoustic model and language model which were constituted by the described method. In the voice recognition equipment which a user can customize by the case where sufficient performance is not attained. In order to make the sound and word which are hard to be recognized recognize better, recognition precision may be able to be raised by making a user do adaptation of the acoustic model, or adding a recognition object word to a user dictionary.
[0015]First, the case where adaptation of the acoustic model is carried out is explained. Drawing 14 is a block diagram showing the composition of conventional voice recognition equipment provided with an acoustic model adaptation means indicated by the above-mentioned literature 1 to carry out adaptation of the acoustic model to the audio signal 100. In drawing 14, a different portion from drawing 12 in order to adapt an acoustic model to the audio signal 100 to input, It is having the initial acoustic model 1003, the acoustic model adaptation means 1004, and the adaptation finishing acoustic model 1401 and that the audio signal 100 is chosen by the acoustic model adaptation means 1004 and the collation means 101.
[0016]Next, operation of the voice recognition equipment shown in drawing 14 is explained. For example using a maximum-aposteriori-probability-estimation method, the acoustic model adaptation means 1004 carries out [sound / for adaptation] adaptation of the initial acoustic model 1003, and obtains the adaptation finishing acoustic model 1401 from the sounds for adaptation (audio signal 100) and the initial acoustic models 1003 collected before actual recognition. The adaptation method of the acoustic model is shown in Chapter 7 of the literature 3. The collation means 101 inputs the audio signal 100, performs speech recognition with reference to the adaptation finishing acoustic model 1401 and the language model 103, and outputs the recognition result 104.
[0017]Next, the case where a user registers a recognition object word into a dictionary is explained. Drawing 15 is a block diagram showing the composition of the conventional voice recognition equipment to which the user dictionary was added indicated by the above-mentioned literature 1. In drawing 15, a different portion from drawing 12 is having the user dictionary 601.
[0018]Drawing 16 is a figure showing the example of composition of the user dictionary 601. The user dictionary 601 is a meeting of the word which the user registered in order to make the word which is not recognized and the word which is hard to be recognized recognize better, and they are a notation of a word, and the list of the words which consist of reading.
[0019]Next, operation of the voice recognition equipment shown in drawing 15 is explained. Drawing 17 is a flow chart which shows the procedure of the conventional speech recognition processing in which the user dictionary 601 was added. In drawing 17, in the language collation processing of step ST1704, a different point from drawing 13 is that the word registered into the user dictionary 601 is added to a recognition object word, in order to refer to the language model 103 and the user dictionary 601.
[0020]The word registered into the user dictionary 601 enables arbitrary word sequences and connection according to suitable conjunctive probability. For example, in the bigram language model determined only with one word which precedes the appearance conditions of a word, Constant value is given to the probability P (wuser|wi) and P of the user dictionary registration word "wuser" inserted into the arbitrary words "wi and wj" (wj|wuser), and reallocation of the probability value is carried out so that the probability of the whole language model may be set to 1. As a result, the user can obtain the recognition result in which the registered word was contained.
[0021]However, in the composition of the voice recognition equipment shown in drawing 14 and drawing 15. In order to raise recognition precision, by creating the audio signal 100 and the adaptation finishing acoustic model 1401, or registering a word into the user dictionary 601, a user needs to customize the acoustic model 102 and the language model 103 himself, and will force a big burden upon a user.
[0022]Since there are a word which had not appeared, and direction for use which was not taken into consideration at the language model construction time even if it is a case where a word is registered to the user dictionary 601, the appearing probability given to those words may be unsuitable. This may become a cause and recognition precision may fall.
[0023]The acoustic model 102 and the language model 103 which were customized as mentioned above are referred to only from the specific collation means 101. For this reason, the other collation means's 101 use of this acoustic model 102 and language model 103 that were customized will reduce recognition precision.
[0024]For this reason, in language processing systems, such as Chinese character kana conversion and a machine translation. The system which eased the user's burden is proposed, for example like JP,10-260960,A by having a function which updates a dictionary automatically via a network about renewal of the user dictionary 601 among the problems shown above. However, in the above-mentioned gazette, it cannot take into consideration treating the model for pattern recognition like a sound, and it cannot be applied to the model about pattern information like the acoustic model 102.
[0025]Also in renewal of the language model 103, if a user's number of registered words increases, since [which gushes and carries out and becomes easy to produce an error] the short word which is not suitable was inserted, the fall of recognition precision will occur easily. Since it does not exist when the word which carried out additional registration to the user dictionary 601 builds the language model 103, or the operating environment is changing, the fall of this recognition precision, For example, it is because the appearing probability which is not suitable for the appearing probability P (wuser|wi) and P of a user registration word (wj|wuser), etc. may be given. As a result, it becomes easy to obtain an unsuitable recognition result, and recognition precision may fall. In order to prevent this, it is necessary to set up the appearing probability of a word appropriately but, and generally it is difficult for the user itself to give an appropriate value.
[0026]
[Problem(s) to be Solved by the Invention]When customizing an acoustic model and a language model in order to raise recognition precision since conventional voice recognition equipment was constituted as mentioned above, the technical problem that a big burden was applied to a user occurred.
[0027]When using except the voice recognition equipment customized for every user, the technical problem that recognition precision fell occurred.
[0028]The technical problem that an acoustic model could not be updated automatically by customization through a network occurred.
[0029]When the registration to a user dictionary increased, the technical problem that recognition precision fell easily occurred.
[0030]By the server side which was made in order that this invention might solve the above technical problems, and was connected to the network. Acquire the newest sound data or language data, and build the acoustic model or language model in the newest state, or, By building the acoustic model or language model corresponding to a user, and updating users' acoustic model and language model via a network. It aims at obtaining the recording medium which recorded the voice recognition system, the voice recognition equipment, the acoustic model managing server, the language model managing server, voice recognition method, and voice recognition program which raise recognition precision without applying a big burden to a user and in which computer reading is possible.
[0031]In all voice recognition equipments by connecting via a network, It aims at obtaining the recording medium which recorded the voice recognition system, the voice recognition equipment, the acoustic model managing server, the language model managing server, voice recognition method, and voice recognition program which can use the customized acoustic model or language model and in which computer reading is possible.
[0032]By using the dictionary and text which are obtained from a user, and the text collected semi-automatically, It aims at obtaining the recording medium which recorded the voice recognition system, the voice recognition equipment, the acoustic model managing server, the language model managing server, voice recognition method, and voice recognition program to which recognition precision cannot fall easily and in which computer reading is possible, even when a user dictionary becomes large.
[0033]
[Means for Solving the Problem]Voice recognition equipment which a voice recognition system concerning this invention inputs an audio signal, performs speech recognition with reference to an acoustic model which asks for probability of an audio acoustical observed value series, and outputs a recognition result, In a thing provided with an acoustic model managing server which is connected with the above-mentioned voice recognition equipment via a network, acquires updated sound data, and builds the above-mentioned acoustic model, The above-mentioned acoustic model which

the above-mentioned acoustic model managing server built is transmitted to the above-mentioned voice recognition equipment, and an acoustic model which the above-mentioned voice recognition equipment refers to in the case of speech recognition is updated by an acoustic model which the above-mentioned acoustic model managing server transmitted.

[0034]A voice recognition system concerning this invention corresponds to a specific condition directed by ID from which an acoustic model managing server acquired and acquired ID which specifies an acoustic model which voice recognition equipment refers to in the case of speech recognition, Updated sound data is read, an acoustic model depending on the above-mentioned specific condition is built, and it transmits to the above-mentioned voice recognition equipment.

[0035]Voice recognition equipment which a voice recognition system concerning this invention inputs an audio signal, performs speech recognition with reference to a language model which searches for appearing probability of a word sequence, and outputs a recognition result, In a thing provided with a language model managing server which is connected with the above-mentioned voice recognition equipment via a network, acquires updated language data, and builds the above-mentioned language model, The above-mentioned language model which the above-mentioned language model managing server built is transmitted to the above-mentioned voice recognition equipment, and a language model which the above-mentioned voice recognition equipment refers to in the case of speech recognition is updated by a language model which the above-mentioned language model managing server transmitted.

[0036]A voice recognition system concerning this invention corresponds to a specific condition directed by ID from which a language model managing server acquired and acquired ID which specifies a language model which voice recognition equipment refers to in the case of speech recognition, Updated language data is read, a language model depending on the above-mentioned specific condition is built, and it transmits to the above-mentioned voice recognition equipment.

[0037]Language data in which a language model managing server read the above-mentioned user dictionary via a network, and a voice recognition system concerning this invention was updated with reference to a user dictionary with which voice recognition equipment registered a word on the occasion of speech recognition, With reference to the read above-mentioned user dictionary, a language model depending on the above-mentioned user dictionary is built, and it transmits to the above-mentioned voice recognition equipment.

[0038]A language model managing server acquires a text which a user of voice recognition equipment used, and a voice recognition system concerning this invention builds a language model depending on the above-mentioned text with reference to updated language data and the acquired above-mentioned text, and transmits to the above-mentioned voice recognition equipment.

[0039]Voice recognition equipment which a voice recognition system concerning this invention inputs an audio signal, performs speech recognition with reference to an acoustic model which asks for probability of an audio acoustical observed value series, and outputs a recognition result, In a thing provided with an acoustic model managing server which is connected with the above-mentioned voice recognition equipment via a network, and has an initial acoustic model before adaptation, The above-mentioned voice recognition equipment acquires voice data for adaptation from an audio signal inputted as ID which specifies the above-mentioned acoustic model, Voice data acquired ID and for adaptation is transmitted to the above-mentioned acoustic model managing server via a network, Using transmitted voice data for adaptation, the above-mentioned acoustic model managing server carries out adaptation of the above-mentioned initial acoustic model, and matches and stores an adaptation finishing acoustic model in above-mentioned ID to which it was transmitted, and. In response to renewal instructions of an acoustic model from the outside, ID which specifies the above-mentioned acoustic model from the above-mentioned voice recognition equipment via a network is received, Choose and read an adaptation finishing acoustic model corresponding to ID which received out of a stored adaptation finishing acoustic model, transmit to the above-mentioned voice recognition equipment via a network, and the above-mentioned voice recognition equipment an acoustic model referred to in the case of speech recognition, It updates by an adaptation finishing acoustic model which the above-mentioned acoustic model managing server transmitted.

[0040]In a thing provided with a collation means which voice recognition equipment concerning this invention inputs an audio signal as an acoustic model which asks for probability of an audio acoustical observed value series, performs speech recognition with reference to the above-mentioned acoustic model, and outputs a recognition result, An acoustic model built with updated sound data from an acoustic model managing server connected via a network is received, and it has an acoustic model update means which updates an acoustic model which the above-mentioned collation means refers to in the case of speech recognition by an acoustic model which received.

[0041]An acoustic model update means voice recognition equipment concerning this invention from an acoustic model managing server connected via a network. An acoustic model depending on a specific condition of an acoustic model which a collation means refers to in the case of speech recognition built with updated sound data is received, and the above-mentioned collation means updates an acoustic model referred to in the case of speech recognition by an acoustic model which received.

[0042]In a thing provided with a collation means which voice recognition equipment concerning this invention inputs an audio signal as a language model which searches for appearing probability of a word sequence, performs speech recognition with reference to the above-mentioned language model, and outputs a recognition result, A language model built with updated language data from a language model managing server connected via a network is received, and it has a language model update means which updates a language model which the above-mentioned collation means refers to in the case of speech recognition by a language model which received.

[0043]A language model update means voice recognition equipment concerning this invention from a language model managing server connected via a network. A language model depending on a specific condition of a language model which a collation means refers to in the case of speech recognition built with updated language data is received, and the above-mentioned collation means updates a language model referred to in the case of speech recognition by a language model which received.

[0044]Voice recognition equipment concerning this invention is provided with a user dictionary which registered a word which a collation means refers to in the case of speech recognition, A language model update means from a language model managing server connected via a network. A language model depending on a user dictionary which was built with updated language data and which the above-mentioned collation means refers to in the case of speech recognition is received, and the above-mentioned collation means updates a language model referred to in the case of speech recognition by a language model which received.

[0045]A language model update means voice recognition equipment concerning this invention from a language model managing server connected via a network. A language model depending on a text which was built with updated language data and which a user who performs speech recognition used is received, and the above-mentioned collation means updates a language model referred to in the case of speech recognition by a language model which received.

[0046]Voice recognition equipment of this invention is characterized by comprising:

An acoustic model which asks for probability of an audio acoustical observed value series.

A collation means which inputs an audio signal, performs speech recognition with reference to the above-mentioned acoustic model, and outputs a recognition result.

An acoustic model ID acquiring means which acquires ID which specifies the above-mentioned acoustic model.

Read ID which the above-mentioned acoustic model ID acquiring means acquired, and voice data for adaptation is acquired from an inputted audio signal, A voice acquisition means for adaptation which transmits voice data read ID and for acquired adaptation to an acoustic model managing server to which it was connected via a network, An acoustic model update means updated by an adaptation finishing acoustic model which received an adaptation finishing acoustic model by which adaptation was carried out with voice data for the above-mentioned adaptation corresponding to above-mentioned ID from the above-mentioned acoustic model managing server, and received an acoustic model which the above-mentioned collation means refers to in the case of speech recognition.

[0047]An acoustic model managing server of this invention is characterized by comprising:
A sound data acquisition means which acquires updated sound data.
An acoustic model construction means which builds an acoustic model which reads updated sound data which the above-mentioned sound data acquisition means acquired in response to renewal instructions of an acoustic model from the outside, and asks for probability of an audio acoustical observed value series.
An acoustic model transmitting means which transmits an acoustic model built by the above-mentioned acoustic model construction means to voice recognition equipment which performs speech recognition via a network.

[0048]An acoustic model managing server of this invention is characterized by comprising:
A sound data acquisition means which acquires updated sound data.
An updating acoustic model ID acquiring means which acquires ID which specifies an acoustic model which voice recognition equipment connected via a network refers to in response to renewal instructions of an acoustic model from the outside in the case of speech recognition.
A sound data reading means for the specification which reads updated sound data which the above-mentioned sound data acquisition means acquired corresponding to a specific condition directed by ID which the above-mentioned updating acoustic model ID acquiring means acquired.
An acoustic model construction means for the specification which builds an acoustic model depending on the above-mentioned specific condition with reference to updated sound data which the above-mentioned sound data reading means for specification read, An acoustic model transmitting means which transmits an acoustic model which the above-mentioned acoustic model construction means for specification built to the above-mentioned voice recognition equipment via a network.

[0049]An acoustic model managing server of this invention is characterized by comprising:
An initial acoustic model before adaptation which asks for probability of an audio acoustical observed value series.
Voice data for adaptation transmitted from voice recognition equipment connected via a network.
An acoustic model adaptation means for the above-mentioned voice recognition equipment to receive ID which specifies an acoustic model referred to in the case of speech recognition, to carry out adaptation of the above-mentioned initial acoustic model using voice data for the above-mentioned adaptation, to match an adaptation finishing acoustic model with above-mentioned ID which received, and to store in an adaptation finishing acoustic model storing means.
In response to renewal instructions of an acoustic model from the outside, above-mentioned ID is received from the above-mentioned voice recognition equipment via a network, An adaptation finishing acoustic model selecting means which chooses and reads an adaptation finishing acoustic model corresponding to ID which received from the above-mentioned adaptation finishing acoustic model storing means, An acoustic model transmitting means which transmits an adaptation finishing acoustic model which the above-mentioned adaptation finishing acoustic model selecting means read to the above-mentioned voice recognition equipment via a network.

[0050]A language model managing server of this invention is characterized by comprising:
A language data acquisition means which acquires updated language data.
A language model construction means which builds a language model which reads updated language data which the above-mentioned language data acquisition means acquired in response to renewal instructions of a language model from the outside, and searches for appearing probability of a word sequence.
A language model transmitting means which transmits a language model which the above-mentioned language model construction means built to voice recognition equipment which performs speech recognition via a network.

[0051]A language model managing server of this invention is characterized by comprising:
A language data acquisition means which acquires updated language data.
An updating language model ID acquiring means which acquires ID which specifies a language model which voice recognition equipment connected via a network refers to in response to renewal instructions of a language model from the outside in the case of speech recognition.
A language data reading means for the specification which reads updated language data which the above-mentioned language data acquisition means acquired corresponding to a specific condition directed by ID which the above-mentioned updating language model ID acquiring means acquired.
A language model construction means for the specification which builds a language model depending on the above-mentioned specific condition with reference to updated language data which the above-mentioned language data reading means for specification read, A language model transmitting means which transmits a language model which the above-mentioned language model construction means for specification built to the above-mentioned voice recognition equipment via a network.

[0052]A language model managing server of this invention is characterized by comprising:
A language data acquisition means which acquires updated language data.
A user dictionary reading means which reads a user dictionary which voice recognition equipment connected via a network refers to in response to renewal instructions of a language model from the outside in the case of speech recognition.
A user dictionary dependence language model construction means which builds a language model depending on a user dictionary which read updated language data which the above-mentioned language data acquisition means acquired, and the above-mentioned user dictionary reading means read.
A language model transmitting means which transmits a language model which the above-mentioned user dictionary dependence language model construction means built to the above-mentioned voice recognition equipment via a network.

[0053]A language model managing server of this invention is characterized by comprising:
A language data acquisition means which acquires updated language data.
A user use text acquisition means which acquires a text which a user of voice recognition equipment connected via a network used in response to renewal instructions of a language model from the outside.
A user use text dependence language model construction means which builds a language model depending on a text which read updated language data which the above-mentioned language data acquisition means acquired, and the above-mentioned user use text acquisition means acquired.
A language model transmitting means which transmits a language model which the above-mentioned user use text dependence language model construction means built to the above-mentioned voice recognition equipment via a network.

[0054]A voice recognition method of this invention inputs an audio signal, performs speech recognition with reference to an acoustic model which asks for probability of an audio acoustical observed value series, and is characterized by that what outputs a recognition result comprises the following.
The 1st step that acquires updated sound data.
The 2nd step that reads updated sound data which was acquired at the 1st step of the above in response to renewal instructions of an acoustic model, and builds an acoustic model.
The 3rd step that transmits an acoustic model built at the 2nd step of the above via a network.
The 4th step updated by an acoustic model which received an acoustic model which transmitted at the 3rd step of the above, and received an acoustic

model referred to in the case of the above-mentioned speech recognition.

[0055]A voice recognition method of this invention inputs an audio signal, performs speech recognition with reference to a language model which searches for appearing probability of a word sequence, and is characterized by that what outputs a recognition result comprises the following.
The 1st step that acquires updated language data.
The 2nd step that reads updated language data which was acquired at the 1st step of the above in response to renewal instructions of a language model, and builds a language model.
The 3rd step that transmits a language model built at the 2nd step of the above via a network.
The 4th step updated by a language model which received a language model which transmitted at the 3rd step of the above, and received a language model referred to in the case of the above-mentioned speech recognition.

[0056]A voice recognition method of this invention inputs an audio signal, performs speech recognition with reference to an acoustic model which asks for probability of an audio acoustical observed value series, and is characterized by that what outputs a recognition result comprises the following.
The 1st step that acquires updated sound data.
The 2nd step that acquires ID which specifies an acoustic model referred to in the case of speech recognition in response to renewal instructions of an acoustic model.
The 3rd step that reads updated sound data which was acquired at the 1st step of the above corresponding to a specific condition directed by ID acquired at the 2nd step of the above.
The 4th step that builds an acoustic model depending on the above-mentioned specific condition with reference to updated sound data which was read at the 3rd step of the above, The 6th step updated by an acoustic model which received an acoustic model which transmitted an acoustic model built at the 4th step of the above at the 5th step that transmits via a network, and the 5th step of the above, and received an acoustic model referred to in the case of speech recognition.

[0057]A voice recognition method of this invention inputs an audio signal, performs speech recognition with reference to a language model which searches for appearing probability of a word sequence, and is characterized by that what outputs a recognition result comprises the following.
The 1st step that acquires updated language data.
The 2nd step that acquires ID which specifies a language model referred to in the case of speech recognition in response to renewal instructions of a language model.
The 3rd step that reads updated language data which was acquired at the 1st step of the above corresponding to a specific condition directed by ID acquired at the 2nd step of the above.
The 4th step that builds a language model depending on the above-mentioned specific condition with reference to updated language data which was read at the 3rd step of the above, The 6th step updated by a language model which received a language model which transmitted a language model built at the 4th step of the above at the 5th step that transmits via a network, and the 5th step of the above, and received a language model referred to in the case of speech recognition.

[0058]A voice recognition method of this invention inputs an audio signal, performs speech recognition with reference to a language model which searches for appearing probability of a word sequence, and a user dictionary which registered a word, and is characterized by that what outputs a recognition result comprises the following.
The 1st step that acquires updated language data.
The 2nd step that reads a user dictionary referred to in the case of speech recognition in response to renewal instructions of a language model.
The 3rd step that builds a language model depending on a user dictionary which read updated language data which was acquired at the 1st step of the above, and was read at the 2nd step of the above.
The 5th step updated by a language model which received a language model which transmitted a language model built at the 3rd step of the above at the 4th step that transmits via a network, and the 4th step of the above, and received a language model referred to in the case of speech recognition.

[0059]A voice recognition method of this invention inputs an audio signal, performs speech recognition with reference to a language model which searches for appearing probability of a word sequence, and is characterized by that what outputs a recognition result comprises the following.
The 1st step that acquires updated language data.
The 2nd step that acquires a text which a user who performs speech recognition used in response to renewal instructions of a language model.
The 3rd step that builds a language model depending on a text which read updated language data which was acquired at the 1st step of the above, and was acquired at the 2nd step of the above.
The 5th step updated by a language model which received a language model which transmitted a language model built at the 3rd step of the above at the 4th step that transmits via a network, and the 4th step of the above, and received a language model referred to in the case of speech recognition.

[0060]A voice recognition method of this invention inputs an audio signal, performs speech recognition with reference to an acoustic model which asks for probability of an audio acoustical observed value series, and is characterized by that what outputs a recognition result comprises the following.
The 1st step that acquires ID which specifies the above-mentioned acoustic model.
The 2nd step that transmits voice data ID which read ID acquired at the 1st step of the above, acquired voice data for adaptation from an inputted audio signal, and was read via a network, and for acquired adaptation.
The 3rd step that carries out adaptation of the initial acoustic model before adaptation, and matches and stores an adaptation finishing acoustic model in ID which transmitted at the 2nd step of the above using voice data for adaptation transmitted at the 2nd step of the above.
In response to renewal instructions of an acoustic model, ID acquired at the 1st step of the above via a network is received, The 4th step that chooses and reads an adaptation finishing acoustic model corresponding to ID which received out of an adaptation finishing acoustic model stored at the 3rd step of the above, The 5th step that transmits an adaptation finishing acoustic model read at the 4th step of the above via a network, The 6th step updated by an adaptation finishing acoustic model which received an adaptation finishing acoustic model which transmitted at the 5th step of the above, and received an acoustic model referred to in the case of speech recognition.

[0061]A recording medium which recorded a voice recognition program concerning this invention and in which computer reading is possible, A sound data acquisition function which acquires sound data which inputted an audio signal, performed speech recognition with reference to an acoustic model which asks for probability of an audio acoustical observed value series, makes realize a verification function which outputs a recognition result, and was updated, An acoustic model construction function to read updated sound data which the above-mentioned sound data acquisition function acquired in response to renewal instructions of an acoustic model, and to build an acoustic model, An acoustic model transmitting function which transmits an acoustic model which the above-mentioned acoustic model construction function built via a network, An acoustic model which the above-mentioned acoustic model transmitting function transmitted is received, and an acoustic model update function which the above-mentioned verification function updates by an acoustic model which received an acoustic model referred to in the case of speech recognition is realized.
[0062]A recording medium which recorded a voice recognition program concerning this invention and in which computer reading is possible, A language

data acquisition function which acquires language data which inputted an audio signal, performed speech recognition with reference to a language model which searches for appearing probability of a word sequence, makes realize a verification function which outputs a recognition result, and was updated, A language model construction function to read updated language data which the above-mentioned language data acquisition function acquired in response to renewal instructions of a language model, and to build a language model, A language model transmitting function which transmits a language model which the above-mentioned language model construction function built via a network, A language model which the above-mentioned language model transmitting function transmitted is received, and a language model update function which the above-mentioned verification function updates by a language model which received a language model referred to in the case of speech recognition is realized.

[0063]A recording medium which recorded a voice recognition program concerning this invention and in which computer reading is possible, A sound data acquisition function which acquires sound data which inputted an audio signal, performed speech recognition with reference to an acoustic model which asks for probability of an audio acoustical observed value series, makes realize a verification function which outputs a recognition result, and was updated, An updating acoustic model ID acquisition function which acquires ID which specifies the above-mentioned acoustic model in response to renewal instructions of an acoustic model, It corresponds to a specific condition directed by ID which the above-mentioned updating acoustic model ID acquisition function acquired, A sound data read-out function for the specification which reads updated sound data which the above-mentioned sound data acquisition function acquired, An acoustic model construction function for the specification which builds an acoustic model depending on the above-mentioned specific condition with reference to updated sound data which the above-mentioned sound data read-out function for specification read, An acoustic model transmitting function which transmits an acoustic model which the above-mentioned acoustic model construction function for specification built via a network, An acoustic model which the above-mentioned acoustic model transmitting function transmitted is received, and an acoustic model update function which the above-mentioned verification function updates by an acoustic model which received an acoustic model referred to in the case of speech recognition is realized.

[0064]A recording medium which recorded a voice recognition program concerning this invention and in which computer reading is possible, A language data acquisition function which acquires language data which inputted an audio signal, performed speech recognition with reference to a language model which searches for appearing probability of a word sequence, makes realize a verification function which outputs a recognition result, and was updated, An updating language model ID acquisition function which acquires ID which specifies the above-mentioned language model in response to renewal instructions of a language model, It corresponds to a specific condition directed by ID which the above-mentioned updating language model ID acquisition function acquired, A language data read-out function for the specification which reads updated language data which the above-mentioned language data acquisition function acquired, A language model construction function for the specification which builds a language model depending on the above-mentioned specific condition with reference to updated language data which the above-mentioned language data read-out function for specification read, A language model transmitting function which transmits a language model which the above-mentioned language model construction function for specification built via a network, A language model which the above-mentioned language model transmitting function transmitted is received, and a language model update function which the above-mentioned verification function updates by a language model which received a language model referred to in the case of speech recognition is realized.

[0065]A recording medium which recorded a voice recognition program concerning this invention and in which computer reading is possible, Input an audio signal and speech recognition is performed with reference to a language model which searches for appearing probability of a word sequence, and a user dictionary which registered a word, A language data acquisition function which acquires language data which makes realize a verification function which outputs a recognition result, and was updated, A user dictionary read-out function which reads the above-mentioned user dictionary in response to renewal instructions of a language model, A user dictionary dependence language model construction function to build a language model depending on a user dictionary which read updated language data which the above-mentioned language data acquisition function acquired, and the above-mentioned user dictionary read-out function read, A language model transmitting function which transmits a language model which the above-mentioned user dictionary dependence language model construction function built via a network, A language model which the above-mentioned language model transmitting function transmitted is received, and a language model update function which the above-mentioned verification function updates by a language model which received a language model referred to in the case of speech recognition is realized.

[0066]A recording medium which recorded a voice recognition program concerning this invention and in which computer reading is possible, A language data acquisition function which acquires language data which inputted an audio signal, performed speech recognition with reference to a language model which searches for appearing probability of a word sequence, makes realize a verification function which outputs a recognition result, and was updated, A user use text acquisition function which acquires a text which a user who performs speech recognition used in response to renewal instructions of a language model, A user use text dependence language model construction function to build a language model depending on a text which read updated language data which the above-mentioned language data acquisition function acquired, and the above-mentioned user use text acquisition function acquired, A language model transmitting function which transmits a language model which the above-mentioned user use text dependence language model construction function built via a network, A language model which the above-mentioned language model transmitting function transmitted is received, and a language model update function which the above-mentioned verification function updates by a language model which received a language model referred to in the case of speech recognition is realized.

[0067]A recording medium which recorded a voice recognition program concerning this invention and in which computer reading is possible, Input an audio signal and speech recognition is performed with reference to an acoustic model which asks for probability of an audio acoustical observed value series, An acoustic model ID acquisition function which acquires ID which realizes a verification function which outputs a recognition result and specifies the above-mentioned acoustic model, Read ID which the above-mentioned acoustic model ID acquisition function acquired, acquire voice data for adaptation from an inputted audio signal, and via a network, A voice acquisition function for adaptation which transmits voice data read ID and for acquired adaptation, Voice data for adaptation which the above-mentioned voice acquisition function for adaptation transmitted is used, An acoustic model adaptation function to carry out adaptation of the initial acoustic model before adaptation, and to match and store an adaptation finishing acoustic model in ID which the above-mentioned voice acquisition function for adaptation transmitted, In response to renewal instructions of an acoustic model, ID which the above-mentioned acoustic model ID acquisition function acquired via a network is received, An adaptation finishing acoustic model function preselection capability which chooses and reads an adaptation finishing acoustic model corresponding to ID which received out of an adaptation finishing acoustic model which the above-mentioned acoustic model adaptation function stored, An acoustic model transmitting function which transmits an adaptation finishing acoustic model which the above-mentioned adaptation finishing acoustic model function preselection capability read via a network, An adaptation finishing acoustic model which the above-mentioned acoustic model transmitting function transmitted is received, and an acoustic model update function which the above-mentioned verification function updates by an adaptation finishing acoustic model which received an acoustic model referred to in the case of speech recognition is realized.

[0068]

[Embodiment of the Invention]Hereafter, one gestalt of implementation of this invention is explained.

Embodiment 1. drawing 1 is a block diagram showing the composition of the voice recognition system by this embodiment of the invention 1. In a figure, the voice recognition equipment in which 10 performs speech recognition, the acoustic model managing server with which 20 is connected to the network, and 30 are language model managing servers connected to the network. Here, a network shows the general communication path which can transmit a digital signal by a cable or radio.

[0069]The audio signal which 100 inputs in the voice recognition equipment 10, the collation means to which 101 carries out speech recognition of the audio signal 100, The acoustic model in which the collation means 101 refers to 102 in the case of speech recognition, the language model in which the collation means 101 refers to 103 in the case of speech recognition, The recognition result to which the collation means 101 outputs 104, the acoustic model update means which updates the acoustic model 102 by the acoustic model to which 111 was transmitted via the network, and 118 are language

model update means which update the language model 103 by the language model transmitted via the network.
[0070]The renewal instructions of an acoustic model to which 105 is given from the exterior in the acoustic model managing server 20, The sound data acquisition means which acquires the sound data in which 106 was updated, the updated sound data in which the sound data acquisition means 106 acquired 107, 108 reads the updated sound data 107 in response to the renewal instructions 105 of an acoustic model, The acoustic model construction means which performs parameter estimation using a statistical method and builds an acoustic model, The acoustic model storing means in which 109 stores the built acoustic model, and 110 are acoustic model transmitting means which transmit the acoustic model stored in the acoustic model storing means 109 to the voice recognition equipment 10 via a network.
[0071]The renewal instructions of a language model to which 112 is given from the exterior in the language model managing server 30, The language data acquisition means which acquires the language data in which 113 was updated, the updated language data in which the language data acquisition means 113 acquired 114, 115 reads the updated language data 114 in response to the renewal instructions 112 of a language model, The language model construction means which performs parameter estimation using a statistical method and builds a language model, The language model storing means in which 116 stores the built language model, and 117 are language model transmitting means which transmit the language model stored in the language model storing means 116 to the voice recognition equipment 10 via a network.
[0072]The characteristic portion of this different invention from conventional technology, In the acoustic model managing server 20 and the language model managing server 30, by the sound data acquisition means 106 and the language data acquisition means 113. The updated sound data 107 and the updated language data 114 are acquired, By the acoustic model construction means 108 and the language model construction means 115, the newest acoustic model, In [build the newest language model, transmit the built newest acoustic model and the newest language model to the voice recognition equipment 10 via a network, and] the voice recognition equipment 10, The acoustic model update means 111 and the language model update means 118 are updating the acoustic model 102 which the collation means 101 refers to, and the language model 103 by the newest acoustic model and the newest language model.
[0073]Next, operation is explained. In the acoustic model managing server 20, it operates asynchronously, and the sound data acquisition means 106 downloads automatically or semi-automatically the renewal instructions 105 of an acoustic model, a synchronization, or the sound data always updated or distributed, and stores it in the updated sound data 107. These sound data to acquire is updated, for example on the Internet, or, It is the voice data distributed by multimedia broadcast or voice data, and a corresponding opening sentence text, and it is searched on the Internet by a retrieving tool, or determines and downloads using the race card of multimedia broadcast.
[0074]The updated sound data 107 is accumulation of the sound data for acoustic model study acquired by the sound data acquisition means 106, and consists of voice data, voice data, and a corresponding opening sentence text in the above-mentioned example.
[0075]A time interval with the acoustic model construction means 108 constant, for example, the time interval with which speech recognition processing was carried out, Or the renewal instructions 105 of an acoustic model given to suitable timing, such as a user's directions given from an input device, are received, With reference to the updated sound data 107, by performing parameter estimation of an acoustic model using a statistical method, For example, by using a vector quantization algorithm only from voice data, Or by using a BAUMU Welch algorithm from voice data and a corresponding opening sentence text, an acoustic model is built and it stores in the acoustic model storing means 109 so that learned data may be expressed well.
[0076]The acoustic model storing means 109 memorizes the acoustic model built by the acoustic model construction means 108, and outputs an acoustic model according to a read request. The acoustic model transmitting means 110 reads an acoustic model from the acoustic model storing means 109, and transmits to the acoustic model update means 111 of the voice recognition equipment 10 via a network.
[0077]In the voice recognition equipment 10, the acoustic model update means 111 updates the acoustic model 102 referred to in the case of collation of the collation means 101 by the acoustic model received from the acoustic model transmitting means 110 of the acoustic model managing server 20 via the network.
[0078]In the language model managing server 30, the language data acquisition means 113, It operates asynchronously and the renewal instructions 112 of a language model, a synchronization, or the language data always updated or distributed is downloaded, and the new language data used for language model construction is collected, and is stored in the updated language data 114. These language data to acquire is a text, texts, such as a chat, mail, and a manual, etc. which can be searched from on the newspaper and mail magazine which are distributed periodically, for example, or the Internet.
[0079]The updated language data 114 is accumulation of the language model study declinable word word data acquired by the language data acquisition means 113, and is text data, the keyword information about the contents of a text acquired simultaneously, etc.
[0080]A time interval with the language model construction means 115 constant, for example, the time interval with which speech recognition processing was carried out, Or the renewal instructions 112 of a language model given to suitable timing, such as a user's directions given from an input device, are received, By reading text data with reference to the updated language data 114, and performing parameter estimation of a language model using a statistical method, For example, by calculating n gram statistics value from the text data divided into the word, a language model is built so that the data for study may be expressed well, and it stores in the language model storing means 116.
[0081]The language model storing means 116 memorizes the language model built by the language model construction means 115, and outputs a language model according to a read request. The language model transmitting means 117 reads a language model from the language model storing means 116, and transmits to the language model update means 118 of the voice recognition equipment 10 via a network.
[0082]In the voice recognition equipment 10, the language model update means 118 updates the language model 103 referred to in the case of collation of the collation means 101 by the language model received from the language model transmitting means 117 of the language model managing server 30 via the network.
[0083]Drawing 2 is a flow chart which shows the update process of the acoustic model 102 by this embodiment of the invention 1. In step ST201, an acoustic model update timing determination means (not shown), For example, suitable update timing is judged from the surveillance of the using state of the time interval from the last modification time of a user's directions or an acoustic model, and a network, etc., and the renewal instructions 105 of an acoustic model are transmitted to the acoustic model construction means 108 of the acoustic model managing server 20. If the renewal instructions 105 of an acoustic model are received, the acoustic model construction means 108 progresses to step ST202, and if it has not received the renewal instructions 105 of an acoustic model, it will end processing.
[0084]In step ST202, the acoustic model construction means 108 reads the updated sound data 107 which is used for study. In step ST203, by performing parameter estimation of an acoustic model using a statistical method, the acoustic model construction means 108 builds an acoustic model, and stores the built acoustic model in the acoustic model storing means 109 from the updated sound data 107.
[0085]In step ST204, the acoustic model transmitting means 110 reads an acoustic model from the acoustic model storing means 109, and transmits an acoustic model to the acoustic model update means 111 of the voice recognition equipment 10 via a network. In step ST205, the acoustic model update means 111 updates the acoustic model 102 which the collation means 101 refers to by the received acoustic model.
[0086]When requiring renewal of the acoustic model 102 by the renewal instructions 105 of an acoustic model, The version of the acoustic model 102 which the voice recognition equipment 10 uses simultaneously is transmitted, If the acoustic model transmitting means 110 transmits only difference information with not the whole acoustic model stored in the acoustic model storing means 109 but the acoustic model 102 which the voice recognition equipment 10 used till then, send data can be reduced and network load can be reduced.
[0087]The acoustic model construction means 108 builds the acoustic model updated beforehand, and even when a gestalt which transmits an acoustic model to the voice recognition equipment 10 is taken according to the demand of the renewal instructions 105 of an acoustic model, it can operate similarly.
[0088]Even if it is a case where the voice recognition equipment 10 has a user dictionary, it can process similarly.

[0089]In this embodiment, although it explains for speech recognition, if it is aimed at the pattern recognition which consists of a stochastic model showing the relation between a pattern and a symbol, and a stochastic model showing the appearance of the symbol, it is applicable similarly.
[0090]As long as the storing form of the updated sound data 107 is a form available at the time of acoustic model construction, it may calculate signal processing and frequency distribution beforehand.
[0091]Drawing 3 is a flow chart which shows the update process of the language model 103 by this embodiment of the invention 1. In step ST301, a language model update timing determination means (not shown), For example, suitable update timing is judged from the surveillance of the using state of the time interval from the last modification time of a user's directions or a language model, and a network, etc., and the renewal instructions 112 of a language model are transmitted to the language model construction means 115 of the language model managing server 30. If the renewal instructions 112 of a language model are received, and the language model construction means 115 progresses to step ST302 and has not received the renewal instructions 112 of a language model, it ends processing.
[0092]In step ST302, the language model construction means 115 reads the updated language data 114 which is used for study. In step ST303, by performing parameter estimation of a language model using a statistical method, the language model construction means 115 builds a language model, and stores the built language model in the language model storing means 116 from the updated language data 114.
[0093]In step ST304, the language model transmitting means 117 reads a language model from the language model storing means 116, and transmits to the language model update means 118 of the voice recognition equipment 10 via a network. In step ST305, the language model update means 118 updates the language model 103 which the collation means 101 refers to by the received language model.
[0094]By transmitting the version of the language model 103 which the voice recognition equipment 10 uses simultaneously by the renewal instructions 112 of a language model when requiring renewal of a language model, If only difference information with not the whole language model stored in the language model storing means 116 but the language model 103 which the voice recognition equipment 10 used till then is transmitted, send data can be reduced and network load can be reduced.
[0095]The language model construction means 115 builds the language model updated beforehand, and even when a gestalt which transmits a language model is taken according to the demand of the renewal instructions 112 of a language model, it can operate similarly.
[0096]Even if it is a case where the voice recognition equipment 10 has a user dictionary, it can operate similarly.
[0097]In this embodiment, although it explains for speech recognition, if it is aimed at the pattern recognition which consists of a stochastic model showing the relation between a pattern and a symbol, and a stochastic model showing the appearance of the symbol, it is applicable similarly.
[0098]The storing form of the updated language data 114, if it is a form available at the time of language model construction, it can divide into the word beforehand or can use for language model construction -- as -- a word, a word chain, the combination of the word which appears simultaneously, etc. -- frequency -- or probability calculation may be carried out.
[0099]Although drawing 1 of this Embodiment 1 shows the case where an acoustic model is built according to the renewal instructions 105 of an acoustic model, When the sound data acquisition means 106 acquires sound data, the acoustic model construction means 108 updates an acoustic model, and it stores in an acoustic model storing means 109, and may be made for the acoustic model transmitting means 110 to read an acoustic model in response to the renewal instructions 105 of an acoustic model. The same may be said of renewal of a language model.
[0100]Although drawing 1 showed the case where it had the acoustic model update means 111 and the language model update means 118, you may be a case where it has either the acoustic model update means 111 or the language model update means 118.
[0101]It is also recordable on a recording medium by making the voice recognition system in Embodiment 1 into a voice recognition program. In this case, the sound data acquisition function which performs the same processing as the sound data acquisition means 106 in the acoustic model managing server 20, The acoustic model construction function to perform the same processing as the acoustic model construction means 108, In the software which comprises an acoustic model storing function to perform the same processing as the acoustic model storing means 109, and an acoustic model transmitting function which performs the same processing as the acoustic model transmitting means 110, and the language model managing server 30, The language data acquisition function which performs the same processing as the language data acquisition means 113, The language model construction function to perform the same processing as the language model construction means 115, In the software which comprises a language model storing function to perform the same processing as the language model storing means 116, and a language model transmitting function which performs the same processing as the language model transmitting means 117, and the voice recognition equipment 10, It becomes a voice recognition program by the software which comprises an acoustic model update function which performs the same processing as the acoustic model update means 111, a language model update function which performs the same processing as the language model update means 118, and a verification function which performs the same processing as the collation means 101.
[0102]The voice recognition program recorded on a recording medium The software of the voice recognition equipment 10, The software of the acoustic model managing server 20 may be recorded on a separate recording medium, may be recorded on one recording medium, and it may transmit to the acoustic model managing server 20 or the voice recognition equipment 10, respectively from the voice recognition equipment 10 or the acoustic model managing server 20. This is the same even when a language model is targeted.
[0103]With as mentioned above, the acoustic model managing server 20 or the language model managing server 30 which was connected to the network according to this Embodiment 1. The updated sound data 107 or the updated language data 114 is acquired, By building the acoustic model or language model in the newest state, and updating users' acoustic model 102 or language model 103 of the voice recognition equipment 10 via a network. The effect that the recognition precision of speech recognition can be raised is acquired without applying a big burden to a user.
[0104]Embodiment 2. drawing 4 is a block diagram showing the composition of the voice recognition system by this embodiment of the invention 2. The updating language model ID acquiring means which acquires ID which specifies the language model 103 of the voice recognition equipment 10 which 401 updates in the language model managing server 30 of drawing 4, and 402, Corresponding to the specific condition directed by ID which updating language model ID acquiring means 401 acquired, the language data reading means for the specification which reads the updated language data 114, and 403, It is a language model construction means for the specification which builds the language model corresponding to a specific condition with reference to the updated language data 114 which was read by the language data reading means 402 for specification. About each means and each model which were already explained, the same numerals are attached and explanation is omitted.
[0105]A portion characteristic of this different embodiment from conventional technology is providing the language model which was provided with updating language model ID acquiring means 401, the language data reading means 402 for specification, and the language model construction means 403 for specification, and was specified by language model ID via the network. Here, the language model for specification is a language model learned so that higher performance might be obtained by making a language model specialize in a specific user, a group, application application, etc.
[0106]Next, operation is explained. In the language managing server 30, updating language model ID acquiring means 401 acquires ID used in order to choose the specific language model 103 from two or more updated language models 103. This ID is a user's user ID, ID showing the task which is the target of the audio signal 100, etc., and can provide the language model 103 for the specification to update in a meaning, for example.
[0107]The language data reading means 402 for specification receives updating language model ID which updating language model ID acquiring means 401 acquired, and reads the updated language data 114 in a sentence or the independent unit of a text, For example, it judges from the keyword about the contents of a text given to language data, or the keyword contained in language data, the flag which identifies whether it is an object specified by language model ID is given, and it sends to the language model construction means 403 for specification.
[0108]The language model construction means 403 for specification builds the language model learned from the updated language data 114 so that recognition precision might become high about a specific object, and stores it in the language model storing means 116.
[0109]For this reason, first study declinable word word data from the keyword etc. which are contained considering a sentence or two or more sentences as a unit. Judge the language model 103 specified and according to this, for example "Examination of the prior task adaptation for interactive

voice recognition", it inquires in Akinori Ito, Masanori Koda, the Institute of Electronics, Information and Communication Engineers technical research report (SP96-81), and 1996 (it is hereafter considered as the literature 4) -- as -- relation -- the language model corresponding to a specific condition can be built by giving big dignity by deep text data.

[0110]For example, in order to build the language model for the specification which specialized in the topic about a sport, With reference to the flag obtained from the language data reading means 402 for specification at the time of language model study, if it is text data of the topic about a sport, actual frequency will be carried out alpha double and will be counted, if it is the other report, both will be added by frequency as they are, and a stochastic model will be presumed. Here, alpha determines that the entropy of the language model to the data which is not used for study among the texts for the specification made into the object of speech recognition serves as the minimum.

[0111]Drawing 5 is a flow chart which shows the update process of the language model 103 by this embodiment of the invention 2. In step ST501, a language model update timing determination means (not shown), For example, suitable update timing is judged from the surveillance of the using state of the time interval from a user's directions or the last modification time, and a network, etc., and the renewal instructions 112 of a language model are transmitted to updating language model ID acquiring means 401 of the language model managing server 30. If the renewal instructions 112 of a language model are received, and updating language model ID acquiring means 401 progresses to step ST502 and has not received the renewal instructions 112 of a language model, it ends processing.

[0112]In step ST502, if it is update timing, by means to specify the user group who is using it, a means to specify a task, etc., updating language model ID acquiring means 401 will acquire ID of the language model 103 of an updating place, and will send it to the language data reading means 402 for specification. In step ST503, the language data reading means 402 for specification, According to language model ID for specification, the updated language data 114 is read in a sentence or the independent unit of a text, the flag which distinguishes whether it is an object specified by language model ID is given, and the updated language data 114 is read.

[0113]In step ST504, according to learning algorithm, the language model construction means 403 for specification presumes a language model from the updated language data 114, and stores the presumed language model in the language model storing means 116. In step ST505, the language model transmitting means 117 transmits the read language model to the language model update means 118 of the voice recognition equipment 10 via a network. In step ST506, the language model update means 118 updates the language model 103 which the collation means 101 refers to by the received language model.

[0114]If the suiting language model can be outputted to language model ID, it is not necessary to build the language model for specification beforehand. For example, only study declinable word word data is created depending on language model ID, and a language model may be built according to a demand.

[0115]Although language model construction for specification according to the literature 4 was made into the example in this embodiment, if it is the method of choosing from two or more language models 103, it is applicable similarly using ID which can determine a language model.

[0116]By transmitting the version of the language model 103 which the voice recognition equipment 10 uses simultaneously in the case of a language model update request by the renewal instructions 112 of a language model, Only the difference information from not the whole language model for the built specification but the present language model 103 can be transmitted, and network load can be reduced.

[0117]Even if it is a case where voice recognition equipment has the user dictionary 601, it can process similarly.

[0118]Although this explanation explained for speech recognition, if aimed at the pattern recognition which consists of a stochastic model showing the relation between a pattern and a symbol, and a stochastic model showing the appearance of the symbol, it is applicable similarly.

[0119]Although the language model for specification was built and the language model 103 is updated by the language model for the built specification in this embodiment, the acoustic model for specification can be built and the acoustic model 102 can also be updated by the acoustic model for the built specification. In drawing 4, instead of the language model managing server 30, instead of an acoustic model managing server and the language data acquisition means 113 In that case, a sound data acquisition means, Instead of the sound data and updating language model ID acquiring means 401 which were updated instead of the updated language data 114, an updating acoustic model ID acquiring means, Instead of the language data reading means 402 for specification, the acoustic model reading means for specification, Instead of the language model construction means 403 for specification, the acoustic model construction means for specification, In [have an acoustic model transmitting means instead of the language model storing means 116 instead of an acoustic model storing means and the language model transmitting means 117, and] the voice recognition equipment 10, It has an acoustic model update means instead of the language model update means 118, and an acoustic model update means should just update the acoustic model 102.

[0120]It is also recordable on a recording medium by making the voice recognition system in Embodiment 2 into a voice recognition program. In this case, the language data acquisition function which performs the same processing as the language data acquisition means 113 in the language model managing server 30, The updating language model ID acquisition function which performs the same processing as updating language model ID acquiring means 401, The language data read-out function for the specification which performs the same processing as the language data reading means 402 for specification, The language model construction function for the specification which performs the same processing as the language model construction means 403 for specification, In the software which comprises a language model storing function to perform the same processing as the language model storing means 116, and a language model transmitting function which performs the same processing as the language model transmitting means 117, and the voice recognition equipment 10, It becomes a voice recognition program by the software which comprises a language model update function which performs the same processing as the language model update means 118, and a verification function which performs the same processing as the collation means 101. This is the same even when an acoustic model is targeted.

[0121]The voice recognition program recorded on a recording medium The software of the voice recognition equipment 10, The software of the language model managing server 30 may be recorded on a separate recording medium, may be recorded on one recording medium, and it may transmit to the language model managing server 30 or the voice recognition equipment 10, respectively from the voice recognition equipment 10 or the language model managing server 30. This is the same even when an acoustic model is targeted.

[0122]With as mentioned above, the language model managing server 30 which was connected to the network according to this Embodiment 2. By acquiring the updated language data 114 and moreover acquiring ID of the language model 103 a user's voice recognition equipment 10. By being in the newest state, building the language model for the specification corresponding to a user moreover, and updating the language model 103 of a user's voice recognition equipment 10 via a network. The effect that the recognition precision of speech recognition can be raised is acquired without applying a big burden to a user.

[0123]By updating the language model customized for specification via a network according to this Embodiment 2, Even when a user uses several different collation means 101, the suitable language model 103 for the utilization time of all the collation means 101 is used, and the effect that high recognition precision can be acquired is acquired.

[0124]Embodiment 3. drawing 6 is a block diagram showing the composition of the voice recognition system by this embodiment of the invention 3. In [in the voice recognition equipment 10 of drawing 6 601 is the user dictionary which registered the word referred to in the case of collation of the collation means 101, and] the language model managing server 30, The user dictionary reading means which reads the user dictionary 601 with which the collation means 101 refers to 602 in response to the renewal instructions 112 of a language model via a network, and 603, It is a user dictionary dependence language model construction means which builds the language model depending on the user dictionary 601 with reference to the updated language data 114 and the user dictionary 601 which the user dictionary reading means 602 read. About each means and each model which were already explained, the same numerals are attached and explanation is omitted.

[0125]A portion characteristic of this different embodiment from conventional technology is having had the user dictionary reading means 602 and the user dictionary dependence language model construction means 603.

[0126]Next, operation is explained. The user dictionary reading means 602 of the language model managing server 30 reads the user dictionary 601 which the collation means 101 of the voice recognition equipment 10 refers to in response to the renewal instructions 112 of a language model via a network. The user dictionary dependence language model construction means 603 builds the language model which is updated by the newest state and customized by the user using the word registered into the user dictionary 601, and the updated language data 114.
[0127]Construction of the language model depending on the user dictionary 601, For example, the text in which the word which exists in the user dictionary 601 is contained out of the updated language data 114 is extracted, and it considers that this is a text for specification, and is carried out by enforcing the method of the literature 4 statement referred to by Embodiment 2. It is expectable that it is not registered by the original language model among the words of user dictionary 601 statement, but become possible to give a statistics value appropriate to the word which appeared in the updated text with the word to which the suitable statistics value was not given, and recognition precision improves by this.
[0128]Drawing 7 is a flow chart which shows the update process of the language model 103 by this embodiment of the invention 3. In step ST701, a language model update timing determination means (not shown), For example, suitable update timing is judged from the surveillance of the using state of the time interval from a user's directions or the last modification time, and a network, etc., and the renewal instructions 112 of a language model are transmitted to the user dictionary reading means 602 of the language model managing server 30. If the renewal instructions 112 of a language model are received, and the user dictionary reading means 602 progresses to step ST702 and has not received the renewal instructions 112 of a language model, it ends processing.
[0129]In step ST702, the user dictionary reading means 602 reads the user dictionary 601 which the collation means 101 refers to via a network. In step ST703, the user dictionary dependence language model construction means 603 reads the language data 114 updated further. In step ST704, the user dictionary dependence language model construction means 603 builds the language model which was dependent on the user dictionary 601 from the user dictionary 601 and the updated language data 114, and stores it in the language model storing means 116.
[0130]In step ST705, the language model transmitting means 117 transmits the language model depending on the user dictionary 601 read from the language model storing means 116 to the language model update means 118 of the voice recognition equipment 10 via a network. In step ST706, the language model update means 118 updates the language model 103 which the collation means 101 refers to by the received language model.
[0131]In this embodiment, although it explains for speech recognition, if it is aimed at the pattern recognition which consists of a stochastic model showing the relation between a pattern and a symbol, and a stochastic model showing the appearance of the symbol, it is applicable similarly.
[0132]It is also recordable on a recording medium by making the voice recognition system in Embodiment 3 into a voice recognition program. In this case, the language data acquisition function which performs the same processing as the language data acquisition means 113 in the language model managing server 30, The user dictionary read-out function to perform the same processing as the user dictionary reading means 602, The user dictionary dependence language model construction function to perform the same processing as the user dictionary dependence language model construction means 603, In the software which comprises a language model storing function to perform the same processing as the language model storing means 116, and a language model transmitting function which performs the same processing as the language model transmitting means 117, and the voice recognition equipment 10, It becomes a voice recognition program by the software which comprises a language model update function which performs the same processing as the language model update means 118, and a verification function which performs the same processing as the collation means 101.
[0133]The voice recognition program recorded on a recording medium The software of the voice recognition equipment 10, The software of the language model managing server 30 may be recorded on a separate recording medium, may be recorded on one recording medium, and it may transmit to the language model managing server 30 or the voice recognition equipment 10, respectively from the voice recognition equipment 10 or the language model managing server 30.
[0134]With as mentioned above, the language model managing server 30 which was connected to the network according to this Embodiment 3. Acquire the updated language data 114 and, moreover, by reading the user dictionary 601 of a user's voice recognition equipment 10. By being in the newest state, building the language model made to reflect in details more about the word moreover registered into the user dictionary 601, and updating the language model 103 of a user's voice recognition equipment 10 via a network. The effect that the recognition precision of speech recognition can be raised is acquired without applying a big burden to a user, even when a user dictionary becomes large.
[0135]Embodiment 4. drawing 8 is a block diagram showing the composition of the voice recognition system by this embodiment of the invention 4. In the language model managing server 30 of drawing 8, 801, The user use text acquisition means which acquires the text which the user used in response to the renewal instructions 112 of a language model, The user use text storing means in which 802 stores the text which the user use text acquisition means 801 acquired, and 803, It is a user use text dependence language model construction means which builds the language model depending on a text with reference to the updated language data 114 and the text stored in the user use text storing means 802. About each means and each model which were already explained, the same numerals are attached and explanation is omitted.
[0136]A portion characteristic of this different embodiment from conventional technology, It has the user use text acquisition means 801, the user use text storing means 802, and the user use text dependence language model construction means 803, It is building a language model according to the text which the user used, and the text which the user used with reference to the language data 114 updated by the newest state.
[0137]Next, operation is explained. The user use text acquisition means 801 of the language model managing server 30 reads the text file which the user referred to or described by scanning the file and directory which the user specified beforehand, in response to the fact that the renewal instructions 112 of a language model. The user use text storing means 802 stores the text collected by the user use text acquisition means 801.
[0138]With reference to a user use text and the updated language data 114, the user use text dependence language model construction means 803 builds a language model so that recognition precision may become high. In construction of the language model using a user use text, the language model of user use text dependence is built by considering that a user use text is a text for specification for example, and enforcing the method of the literature 4 statement referred to by Embodiment 2. thus, since the user is making the character of the text which carried out reference or existing appearance reflect, the built language model can obtain a higher-precision recognition result including the linguistic character in which the probability which a user utters is high.
[0139]Drawing 9 is a flow chart which shows the update process of the language model 103 by this embodiment of the invention 4. In step ST901, a language model update timing determination means (not shown), Suitable timing is judged from the monitor of the using state of the time interval from a user's directions or the last modification time, and a network, etc., and the renewal instructions 112 of a language model are transmitted to the user use text acquisition means 801 of the language model managing server 30. If the renewal instructions 112 of a language model are received, and the user use text acquisition means 801 progresses to step ST902 and has not received the renewal instructions 112 of a language model, it ends processing.
[0140]In step ST902, the user use text acquisition means 801 reads a user use text, and stores it in the user use text storing means 802. In step ST903, the user use text dependence language model construction means 803 reads a user use text and the updated language data 114. In step ST904, the user use text dependence language model construction means 803 builds a user use text dependence language model from a user use text and the updated language data 114, and stores it in the language model storing means 116.
[0141]In step ST905, the language model transmitting means 117 transmits the user use text dependence language model read from the language model storing means 116 to the language model update means 118 of the voice recognition equipment 10 via a network. In step ST906, the language model update means 118 updates the language model 103 which the collation means 101 refers to by the received language model.
[0142]Although it presupposed that acquisition of a user use text is based on search of a specific directory or a file in this embodiment, The user use text acquisition means 801 will pick out a text from neither a file nor a directory, if a text is collectable, The text etc. which the user perused by user inputs, such as speech recognition, a keyboard, a pen, and OCR, or a browser may be used.
[0143]Although the user use text storing means 802 presupposed that the text collected by the user use text acquisition means 801 is stored, It is the

same, even if a suitable means divides a text into a word or it stores in accordance with the standard of the user use text dependence language model construction means 803 as frequency about the word referred to at the time of modern construction, a word chain, the combination of the word which appears simultaneously, etc.

[0144]In this embodiment, although explained for speech recognition, if aimed at the pattern recognition which consists of a stochastic model showing the relation between a pattern and a symbol, and a stochastic model showing the appearance of the symbol, it is applicable similarly.

[0145]The voice recognition system in Embodiment 4 is also recordable on a recording medium as a voice recognition program. In this case, the language data acquisition function which performs the same processing as a language data acquisition means in the language model managing server 30, The user use text acquisition function which performs the same processing as the user use text acquisition means 801, The user use text storing function to perform the same processing as the user use text storing means 802, The user use text dependence language model construction function to perform the same processing as the user use text dependence language model construction means 803, In language model storing which performs the same processing as the language model storing means 116, the software which comprises a language model transmitting function which performs the same processing as the language model transmitting means 117, and the voice recognition equipment 10, It becomes a voice recognition program by the software which comprises a language model update function which performs the same processing as the language model update means 118, and a verification function which performs the same processing as the collation means 101.

[0146]The voice recognition program recorded on a recording medium The software of the voice recognition equipment 10, The software of the language model managing server 30 may be recorded on a separate recording medium, may be recorded on one recording medium, and it may transmit to the language model managing server 30 or the voice recognition equipment 10, respectively from the voice recognition equipment 10 or the language model managing server 30.

[0147]With as mentioned above, the language model managing server 30 which was connected to the network according to this Embodiment 4. By acquiring the updated language data 114 and moreover acquiring the text which the user used. By being in the newest state, building the language model depending on the text which the user moreover used, and updating the language model 103 of a user's voice recognition equipment 10 via a network. The effect that the recognition precision of speech recognition can be raised is acquired without applying a big burden to a user.

[0148]Embodiment 5. drawing 10 is a block diagram showing the composition of the voice recognition system by this embodiment of the invention 5. The acoustic model ID acquiring means which acquires ID which identifies the acoustic model which refers to 1001 in the case of collation of the collation means 101 in the voice recognition equipment 10 of drawing 10, and 1002, It is a voice acquisition means for adaptation which reads ID which acoustic model ID acquiring means 1001 acquired, acquires the voice data for adaptation from the inputted audio signal, and transmits read ID and the acquired voice data for adaptation to the acoustic model managing server 20 via a network.

[0149]1003 in the acoustic model managing server 20 of drawing 10 the initial acoustic model before adaptation and 1004, The voice data for adaptation transmitted from the voice acquisition means 1002 for adaptation of the voice recognition equipment 10 is used, Carry out adaptation of the initial acoustic model 1003 before adaptation, and an adaptation finishing acoustic model, An acoustic model adaptation means to match with ID transmitted from the voice acquisition means 1002 for adaptation, and to store in the adaptation finishing acoustic model storing means 1005, and 1006, In response to the renewal instructions 105 of an acoustic model, ID which acoustic model ID acquiring means 1001 of the voice recognition equipment 10 acquired via the network is received, It is an adaptation finishing acoustic model selecting means which chooses and reads the adaptation finishing acoustic model corresponding to ID which received from the adaptation finishing acoustic model storing means 1005. About each means and each model which were already explained, the same numerals are attached and explanation is omitted.

[0150]A portion characteristic of this different embodiment from conventional technology, In [in the voice recognition equipment 10, have acoustic model ID acquiring means 1001 and the voice acquisition means 1002 for adaptation, and] the acoustic model managing server 20, It is having had the initial acoustic model 1003, the acoustic model adaptation means 1004, the adaptation finishing acoustic model storing means 1005, and the adaptation finishing acoustic model selecting means 1006.

[0151]In [in this embodiment, acquire the voice data acoustic model ID used as an adaptation object, and for adaptation in the voice recognition equipment 10, and] the acoustic model managing server 20, In [build the acoustic model which performed adaptation depending on acoustic model ID, transmit to the voice recognition equipment 10, and] the voice recognition equipment 10, Since the user can refer to the acoustic model which carried out adaptation when using the arbitrary collation means 101 by updating the acoustic model 102, higher recognition precision can be acquired.

[0152]Next, operation is explained. In the voice recognition equipment 10, acoustic model ID acquiring means 1001 is the user ID of the user who determines the acoustic model used as an adaptation object, and uses the voice recognition equipment 10. Acoustic model ID which the voice acquisition means 1002 for adaptation acquires the voice data for adaptation based on the audio signal 100 as an object for adaptation beforehand before use of the collation means 101, and is read from acoustic model ID acquiring means 1001, The acquired bottom transmits the voice data for adaptation to an acoustic model adaptation means 1004 of the acoustic model managing server 20 by which it is connected via a network.

[0153]In the acoustic model managing server 20, the initial acoustic model 1003 is an acoustic model before performing adaptation. The voice data and the initial acoustic model 1003 for adaptation which were received via the network are used for the acoustic model adaptation means 1004, The acoustic model which carried out adaptation is built and acoustic model ID received via an acoustic model and a network is stored in the adaptation finishing acoustic model storing means 1005. [finishing / adaptation] A maximum-aposteriori-probability-estimation method is used for the adaptation of an acoustic model, for example.

[0154]The adaptation finishing acoustic model storing means 1005 stores the acoustic model by which adaptation was carried out to acoustic model ID by the acoustic model adaptation means 1004, and outputs an acoustic model with appointed acoustic model ID according to a demand of an adaptation finishing acoustic model selecting means. In response to the renewal instructions 105 of an acoustic model, the adaptation finishing acoustic model selecting means 1006 acquires acoustic model ID from acoustic model ID acquiring means 1001, and chooses and reads a corresponding adaptation finishing acoustic model from the adaptation finishing acoustic model storing means 1005.

[0155]Drawing 11 is a flow chart which shows the update process of the acoustic model 102 by this embodiment of the invention 5. Processing is divided into the adaptation stage of the acoustic models from step ST1101 to ST1107, and the updating stage of the acoustic models from step ST1108 to ST1112. In an adaptation stage, the acoustic model depending on acoustic model ID is built using the inputted voice data for adaptation.

[0156]In step ST1101, acoustic model ID acquiring means 1001 of the voice recognition equipment 10 identifies the acoustic model used as an adaptation object, for example, acquires identification information, such as a user name. In step ST1102, the voice acquisition means 1002 for adaptation reads acoustic model ID, and acquires the voice data for adaptation inputted from the audio signal 100. In step ST1103, the voice acquisition means 1002 for adaptation transmits the adaptation demand of an acoustic model to the acoustic model managing server 20 via a network, and transmits the voice data for acoustic model ID and adaptation to the acoustic model adaptation means 1004 simultaneously.

[0157]In step ST1104, the acoustic model adaptation means 1004 receives the adaptation demand of an acoustic model via a network, and reads the voice data for acoustic model ID and adaptation. In step ST1105, the acoustic model adaptation means 1004 reads the initial acoustic model 1003. In step ST1106, the acoustic model adaptation means 1004 carries out adaptation of the initial acoustic model 1003 using the voice data for adaptation received via the network. In step ST1107, the acoustic model adaptation means 1004 is stored in the adaptation finishing acoustic model storing means 1005 so that the acoustic model by which adaptation was carried out can be distinguished by acoustic model ID.

[0158]In step ST1108, an acoustic model update timing determination means (not shown), Suitable update timing is judged from the surveillance of the using state of the time interval from a user's directions or the last modification time, and a network, etc., and the renewal instructions 105 of an acoustic model are transmitted to the adaptation finishing acoustic model selecting means 1006. If the renewal instructions 105 of an acoustic model are received, the adaptation finishing acoustic model selecting means 1006 progresses to step ST1109, and if it has not received the renewal instructions 105 of an acoustic model, it will end processing. In step ST1109, the adaptation finishing acoustic model selecting means 1006 reads

acoustic model ID for adaptation from acoustic model ID acquiring means 1001 of the voice recognition equipment 10 via a network.
[0159]In step ST1110, the adaptation finishing acoustic model selecting means 1006 chooses and reads the acoustic model specified by acoustic model ID from the adaptation finishing acoustic model storing means 1005. In step ST1111, the acoustic model transmitting means 110 transmits the read adaptation finishing acoustic model to the acoustic model update means 111 of the voice recognition equipment 10 via a network. In step ST1112, the acoustic model update means 111 updates the acoustic model 102 which the collation means 101 refers to by the received adaptation finishing acoustic model.
[0160]In the voice recognition equipment 10, as long as acoustic model ID acquiring means 1001 determines the acoustic model used as an adaptation object, it may determine a circuit transfer characteristic, the background-noise characteristic, the reverberant sound characteristic, etc.
[0161]In the voice recognition equipment 10, although the voice acquisition means 1002 for adaptation acquires a user's voice data as an object for adaptation beforehand before use of the collation means 101, It is also possible to perform adaptation of the acoustic model referred to in the case of the next collation in the collation means 101 acquiring and carrying out voice data at the time of collation, and inputting into the voice acquisition means 1002 for adaptation.
[0162]In the voice recognition equipment 10, the procedure of acquisition of the data of the sound for adaptation and acquisition of acoustic model ID may be reverse.
[0163]In the voice recognition equipment 10, although the storing gestalt of the voice data of the voice acquisition means 1002 for adaptation was made into the voice waveform, The time series of the sound feature vector which carried out signal processing of the voice data, the code book code sequence acquired with reference to a sound feature vector and a vector quantization code book, As long as the frequency distribution acquired by carrying out the statistical work of them, the probability distributions acquired from frequency distribution, etc. are the gestalten which can be used for study of an acoustic model, what kind of thing may be used.
[0164]A means to store the voice data for adaptation which the acoustic model adaptation means 1004 received in the acoustic model managing server 20 is added, By updating the early model 1003 with the stored voice data for much adaptation, the accuracy of study becomes high and the adaptation finishing acoustic model which performs highly precise recognition can be built.
[0165]Although speech recognition was targeted in this embodiment, if aimed at the pattern recognition which consists of a stochastic model showing the relation between a pattern and a symbol, and a stochastic model showing the appearance of the symbol, it is applicable similarly.
[0166]It is also recordable on a recording medium by making the voice recognition system in Embodiment 5 into a voice recognition program. In this case, the acoustic model ID acquisition function which performs the same processing as acoustic model ID acquiring means 1001 in the voice recognition equipment 10, The voice acquisition function for adaptation which performs the same processing as the voice acquisition means 1002 for adaptation, In the software which comprises an acoustic model update function which performs the same processing as the acoustic model update means 111, and a verification function which performs the same processing as the collation means 101, and the acoustic model managing server 20, The acoustic model adaptation function to perform the same processing as the acoustic model adaptation means 1004, The adaptation finishing acoustic model storing function to perform the same processing as the adaptation finishing acoustic model storing means 1005, It becomes a voice recognition program by the software which comprises an adaptation finishing acoustic model function preselection capability which performs the same processing as the adaptation finishing acoustic model selecting means 1006, and an acoustic model transmitting function which performs the same processing as the acoustic model transmitting means 110.
[0167]The voice recognition program recorded on a recording medium The software of the voice recognition equipment 10, The software of the acoustic model managing server 20 may be recorded on a separate recording medium, may be recorded on one recording medium, and it may transmit to the acoustic model managing server 20 or the voice recognition equipment 10, respectively from the voice recognition equipment 10 or the acoustic model managing server 20.
[0168]With as mentioned above, the sound language model managing server 20 which was connected to the network according to this Embodiment 5. By building an acoustic model [finishing / adaptation] with the voice data for adaptation based on a user's audio signal 100, and updating the acoustic model 102 of a user's voice recognition equipment 10 via a network. The effect that the recognition precision of speech recognition can be raised is acquired without applying a big burden to a user.
[0169]By what the acoustic model 102 is updated for via a network by the adaptation finishing acoustic model which carried out adaptation to the user according to this Embodiment 5. Even when a user uses several different collation means 101, the suitable acoustic model 102 for the utilization time of all the collation means 101 is used, and the effect that high recognition precision can be acquired is acquired.
[0170]
[Effect of the Invention]As mentioned above, the acoustic model in which the acoustic model managing server acquired and built the updated sound data according to this invention, By transmitting to voice recognition equipment via a network, and updating the acoustic model which voice recognition equipment refers to in the case of speech recognition by the acoustic model which the acoustic model managing server transmitted, It is effective in the ability to raise the recognition precision of speech recognition, without applying a big burden to a user.
[0171]According to this invention, it corresponds to the specific condition directed by ID from which the acoustic model managing server acquired and acquired ID which specifies the acoustic model which voice recognition equipment refers to in the case of speech recognition, Can raise the recognition precision of speech recognition, without applying a big burden to a user by reading the updated sound data, building the acoustic model depending on a specific condition, and transmitting to voice recognition equipment, and. By transmitting the acoustic model customized for specification via a network, even when a user uses several different voice recognition equipments, a suitable acoustic model is used, and it is effective in the ability to acquire high recognition precision.
[0172]According to this invention, a language model managing server the language model which acquired and built the updated language data, By transmitting to voice recognition equipment via a network, and updating the language model which voice recognition equipment refers to in the case of speech recognition by the language model which the language model managing server transmitted, It is effective in the ability to raise the recognition precision of speech recognition, without applying a big burden to a user.
[0173]According to this invention, it corresponds to the specific condition directed by ID from which the language model managing server acquired and acquired ID which specifies the language model which voice recognition equipment refers to in the case of speech recognition, Can raise the recognition precision of speech recognition, without applying a big burden to a user by reading the updated language data, building the language model depending on a specific condition, and transmitting to voice recognition equipment, and. By updating the language model customized for specification via a network, even when a user uses several different voice recognition equipments, a suitable language model is used, and it is effective in the ability to acquire high recognition precision.
[0174]The language data in which the language model managing server read the user dictionary, and was updated via the network with reference to the user dictionary with which voice recognition equipment registered the word on the occasion of speech recognition according to this invention, It is effective in the ability to raise the recognition precision of speech recognition, without applying a big burden to a user, even when a user dictionary becomes large by building the language model depending on a user dictionary with reference to the read user dictionary, and transmitting to voice recognition equipment.
[0175]The language data which according to this invention the language model managing server acquired the text which the user of voice recognition equipment used, and was updated, It is effective in the ability to raise the recognition precision of speech recognition, without applying a big burden to a user by building the language model depending on a text with reference to the acquired text, and transmitting to the above-mentioned voice recognition equipment.
[0176]ID as which voice recognition equipment specifies an acoustic model according to this invention, Acquire the voice data for adaptation from the

inputted audio signal, and the voice data acquired ID and for adaptation, Transmit to an acoustic model managing server via a network, and an acoustic model managing server, Using the transmitted voice data for adaptation, carry out adaptation of the initial acoustic model, and match and store an adaptation finishing acoustic model in ID to which it was transmitted, and. In response to the renewal instructions of an acoustic model from the outside, ID is received from voice recognition equipment via a network, The adaptation finishing acoustic model corresponding to ID which received is chosen and read out of the stored adaptation finishing acoustic model, By transmitting to voice recognition equipment via a network, and updating the acoustic model which voice recognition equipment refers to in the case of speech recognition by the adaptation finishing acoustic model which the acoustic model managing server transmitted, Can raise the recognition precision of speech recognition, without applying a big burden to a user, and. Even when a user uses several different voice recognition equipments by updating an acoustic model via a network by the adaptation finishing acoustic model which carried out adaptation to the user, a suitable acoustic model is used, It is effective in the ability to acquire high recognition precision.

[0177]According to this invention, from the acoustic model managing server to which voice recognition equipment was connected via the network. When the acoustic model built with the updated sound data was received and the collation means was provided with the acoustic model update means which updates the acoustic model referred to in the case of speech recognition by the acoustic model which received, It is effective in the ability to raise the recognition precision of speech recognition, without applying a big burden to a user.

[0178]According to this invention, an acoustic model update means from the acoustic model managing server connected via the network. The acoustic model depending on the specific condition of the acoustic model which a collation means refers to in the case of speech recognition built with the updated sound data is received, Can raise the recognition precision of speech recognition, without applying a big burden to a user, when a collation means updates the acoustic model referred to in the case of speech recognition by the acoustic model which received, and. By receiving the acoustic model customized for specification via a network, even when a user uses several different collation means, a suitable acoustic model is used, and it is effective in the ability to acquire high recognition precision.

[0179]According to this invention, from the language model managing server to which voice recognition equipment was connected via the network. When the language model built with the updated language data was received and the collation means was provided with the language model update means which updates the language model referred to in the case of speech recognition by the language model which received, It is effective in the ability to raise the recognition precision of speech recognition, without applying a big burden to a user.

[0180]According to this invention, a language model update means from the language model managing server connected via the network. The language model depending on the specific condition of the language model which a collation means refers to in the case of speech recognition built with the updated language data is received, Can raise the recognition precision of speech recognition, without applying a big burden to a user, when a collation means updates the language model referred to in the case of speech recognition by the language model which received, and. By receiving the language model customized for specification via a network, even when a user uses several different collation means, a suitable language model is used, and it is effective in the ability to acquire high recognition precision.

[0181]According to this invention, a collation means is provided with the user dictionary which registered the word referred to in the case of speech recognition, A language model update means from the language model managing server connected via the network. When the language model depending on a user dictionary built with the updated language data is received and a collation means updates the language model referred to in the case of speech recognition by the language model which received, It is effective in the ability to raise the recognition precision of speech recognition, without applying a big burden to a user, even when a user dictionary becomes large.

[0182]According to this invention, a language model update means from the language model managing server connected via the network. When the language model depending on the text which was built with the updated language data and which the user who performs speech recognition used is received and a collation means updates the language model referred to in the case of speech recognition by the language model which received, It is effective in the ability to raise the recognition precision of speech recognition, without applying a big burden to a user.

[0183]The acoustic model which asks for the probability of an audio acoustical observed value series according to this invention, The collation means which inputs an audio signal, performs speech recognition with reference to the above-mentioned acoustic model, and outputs a recognition result, The acoustic model ID acquiring means which acquires ID which specifies an acoustic model, and acquired ID are read, The voice acquisition means for adaptation which acquires the voice data for adaptation from the inputted audio signal, and transmits the voice data read ID and for the acquired adaptation to the acoustic model managing server to which it was connected via the network, From an acoustic model managing server, the adaptation finishing acoustic model by which adaptation was carried out with the voice data for adaptation corresponding to ID is received, When the collation means was provided with the acoustic model update means which updates the acoustic model referred to in the case of speech recognition by the adaptation finishing acoustic model which received, Can raise the recognition precision of speech recognition, without applying a big burden to a user, and. It is effective in the ability to use a suitable acoustic model and acquire high recognition precision by updating an acoustic model via a network, by the adaptation finishing acoustic model which carried out adaptation to the user, even when a user uses several different collation means.

[0184]The sound data acquisition means which acquires the updated sound data according to this invention, The acoustic model construction means which builds the acoustic model which reads the sound data updated in response to the renewal instructions of an acoustic model from the outside, and asks for the probability of an audio acoustical observed value series, It is effective in the ability to raise the recognition precision of speech recognition, without applying a big burden to a user by having had the acoustic model transmitting means which transmits the acoustic model built by the acoustic model construction means to the voice recognition equipment which performs speech recognition via a network.

[0185]The sound data acquisition means which acquires the updated sound data according to this invention, The updating acoustic model ID acquiring means which acquires ID which specifies the acoustic model referred to in the case of speech recognition in response to the renewal instructions of an acoustic model from the outside, The sound data reading means for the specification which reads the updated sound data corresponding to the specific condition directed by acquired ID, The acoustic model construction means for the specification which builds the acoustic model depending on a specific condition with reference to the read sound data which was updated, Can raise the recognition precision of speech recognition, without applying a big burden to a user by having had the acoustic model transmitting means which transmits the built acoustic model to voice recognition equipment via a network, and. By transmitting the acoustic model customized for specification via a network, even when a user uses several different voice recognition equipments, a suitable acoustic model is used, and it is effective in the ability to acquire high recognition precision.

[0186]The initial acoustic model before adaptation which asks for the probability of an audio acoustical observed value series according to this invention, The voice data for adaptation transmitted from the voice recognition equipment connected via the network, ID which specifies the acoustic model which voice recognition equipment refers to in the case of speech recognition is received, An acoustic model adaptation means to carry out adaptation of the initial acoustic model using the voice data for adaptation, to match an adaptation finishing acoustic model with ID which received, and to store in an adaptation finishing acoustic model storing means, In response to the renewal instructions of an acoustic model from the outside, ID is received from voice recognition equipment via a network, The adaptation finishing acoustic model selecting means which chooses and reads the adaptation finishing acoustic model corresponding to ID which received from an adaptation finishing acoustic model storing means, Can raise the recognition precision of speech recognition, without applying a big burden to a user by having had the acoustic model transmitting means which transmits the read adaptation finishing acoustic model to voice recognition equipment via a network, and. Even when a user uses several different voice recognition equipments by updating an acoustic model via a network by the adaptation finishing acoustic model which carried out adaptation to the user, a suitable acoustic model is used, It is effective in the ability to acquire high recognition precision.

[0187]The language data acquisition means which acquires the updated language data according to this invention, The language model construction means which builds the language model which reads the language data updated in response to the renewal instructions of a language model from the outside, and searches for the appearing probability of a word sequence, It is effective in the ability to raise the recognition precision of speech recognition, without applying a big burden to a user by having had the language model transmitting means which transmits the language model which the

language model construction means built to the voice recognition equipment which performs speech recognition via a network.
[0188]The language data acquisition means which acquires the updated language data according to this invention, The updating language model ID acquiring means which acquires ID which specifies the language model referred to in the case of speech recognition in response to the renewal instructions of a language model from the outside, The language data reading means for the specification which reads the updated language data corresponding to the specific condition directed by ID, The language model construction means for the specification which builds the language model depending on a specific condition with reference to the read language data which was updated, Can raise the recognition precision of speech recognition, without applying a big burden to a user by having had the language model transmitting means which transmits the built language model to voice recognition equipment via a network, and. By transmitting the language model customized for specification via a network, even when a user uses several different voice recognition equipments, a suitable language model is used, and it is effective in the ability to acquire high recognition precision.
[0189]The language data acquisition means which acquires the updated language data according to this invention, The user dictionary reading means which reads the user dictionary which the voice recognition equipment connected via the network refers to in response to the renewal instructions of a language model from the outside in the case of speech recognition, The user dictionary dependence language model construction means which builds the language model which read the updated language data and was dependent on the user dictionary, It is effective in the ability to raise the recognition precision of speech recognition, without applying a big burden to a user, even when a user dictionary becomes large by having had the language model transmitting means which transmits the built language model to voice recognition equipment via a network.
[0190]The language data acquisition means which acquires the updated language data according to this invention, The user use text acquisition means which acquires the text which the user used in response to the renewal instructions of a language model from the outside, The user use text dependence language model construction means which builds the language model which read the updated language data and depended for it on the acquired text, It is effective in the ability to raise the recognition precision of speech recognition, without applying a big burden to a user by having had the language model transmitting means which transmits the built language model to voice recognition equipment via a network.
[0191]The 1st step that acquires the updated sound data according to this invention, The 2nd step that reads the sound data updated in response to the renewal instructions of an acoustic model, and builds an acoustic model, By having had the 3rd step that transmits the built acoustic model via a network, and the 4th step updated by the acoustic model which received the acoustic model and received the acoustic model referred to in the case of speech recognition, It is effective in the ability to raise the recognition precision of speech recognition, without applying a big burden to a user.
[0192]The 1st step that acquires the updated language data according to this invention, The 2nd step that reads the language data updated in response to the renewal instructions of a language model, and builds a language model, By having had the 3rd step that transmits the built language model via a network, and the 4th step updated by the language model which received the language model and received the language model referred to in the case of speech recognition, It is effective in the ability to raise the recognition precision of speech recognition, without applying a big burden to a user.
[0193]The 1st step that acquires the updated sound data according to this invention, The 2nd step that acquires ID which specifies the acoustic model referred to in the case of speech recognition in response to the renewal instructions of an acoustic model, The 3rd step that reads the updated sound data corresponding to the specific condition directed by ID, The 4th step that builds the acoustic model depending on a specific condition with reference to the updated sound data, By having had the 5th step that transmits the built acoustic model via a network, and the 6th step updated by the acoustic model which received the acoustic model and received the acoustic model referred to in the case of speech recognition, Can raise the recognition precision of speech recognition, without applying a big burden to a user, and. By receiving the acoustic model customized for specification via a network, even when a user uses several different voice recognition methods, a suitable acoustic model is used, and it is effective in the ability to acquire high recognition precision.
[0194]The 1st step that acquires the updated language data according to this invention, The 2nd step that acquires ID which specifies the language model referred to in the case of speech recognition in response to the renewal instructions of a language model, The 3rd step that reads the updated language data corresponding to the specific condition directed by acquired ID, The 4th step that builds the language model depending on a specific condition with reference to the updated language data, By having had the 5th step that transmits the built language model via a network, and the 6th step updated by the language model which received the language model and received the language model referred to in the case of speech recognition, Can raise the recognition precision of speech recognition, without applying a big burden to a user, and. By receiving the language model customized for specification via a network, even when a user uses several different voice recognition methods, a suitable language model is used, and it is effective in the ability to acquire high recognition precision.
[0195]The 1st step that acquires the updated language data according to this invention, The 2nd step that reads the user dictionary referred to in the case of speech recognition in response to the renewal instructions of a language model, The 3rd step that builds the language model which read the updated language data and was dependent on the user dictionary, By having had the 4th step that transmits the built language model via a network, and the 5th step updated by the language model which received the language model and received the language model referred to in the case of speech recognition, It is effective in the ability to raise the recognition precision of speech recognition, without applying a big burden to a user, even when a user dictionary becomes large.
[0196]The 1st step that acquires the updated language data according to this invention, The 2nd step that acquires the text which the user used in response to the renewal instructions of a language model, The 3rd step that builds the language model which read the updated language data and was dependent on the text, By having had the 4th step that transmits the built language model via a network, and the 5th step updated by the language model which received the language model and received the language model referred to in the case of speech recognition, It is effective in the ability to raise the recognition precision of speech recognition, without applying a big burden to a user.
[0197]The 1st step that acquires ID which specifies an acoustic model according to this invention, Read acquired ID, acquire the voice data for adaptation from the inputted audio signal, and via a network, The 2nd step that transmits the voice data read ID and for the acquired adaptation, The 3rd step that carries out adaptation of the initial acoustic model before adaptation, and matches and stores an adaptation finishing acoustic model in ID which transmitted using the transmitted voice data for adaptation, In response to the renewal instructions of an acoustic model, ID acquired at the 1st step via the network is received, The 4th step that chooses and reads the adaptation finishing acoustic model corresponding to ID which received out of the adaptation finishing acoustic model stored at the 3rd step, The 5th step that transmits the adaptation finishing acoustic model read at the 4th step via a network, By having had the 6th step updated by the adaptation finishing acoustic model which received the adaptation finishing acoustic model which transmitted and received the acoustic model referred to in the case of speech recognition, Can raise the recognition precision of speech recognition, without applying a big burden to a user, and by the adaptation finishing acoustic model which carried out adaptation to the user. Even when a user uses several different voice recognition methods by updating an acoustic model via a network, a suitable acoustic model is used and it is effective in the ability to acquire high recognition precision.
[0198]The sound data acquisition function which acquires the updated sound data with the recording medium which recorded the voice recognition program according to this invention, The acoustic model construction function to read the sound data updated in response to the renewal instructions of an acoustic model, and to build an acoustic model, The acoustic model transmitting function which transmits the built acoustic model via a network, It is effective in the ability to raise the recognition precision of speech recognition, without applying a big burden to a user by receiving an acoustic model and realizing the acoustic model update function which a verification function updates by the acoustic model which received the acoustic model referred to in the case of speech recognition.
[0199]The language data acquisition function which acquires the updated language data with the recording medium which recorded the voice recognition program according to this invention, The language model construction function to read the updated language data which the above-mentioned language data acquisition function acquired in response to the renewal instructions of a language model, and to build a language model, The language model transmitting function which transmits the language model which the above-mentioned language model construction function built via a network, By

receiving the language model which the above-mentioned language model transmitting function transmitted, and realizing the language model update function which the above-mentioned verification function updates by the language model which received the language model referred to in the case of speech recognition, It is effective in the ability to raise the recognition precision of speech recognition, without applying a big burden to a user.

[0200]The sound data acquisition function which acquires the updated sound data with the recording medium which recorded the voice recognition program according to this invention, The updating acoustic model ID acquisition function which acquires ID which specifies an acoustic model in response to the renewal instructions of an acoustic model, The sound data read-out function for the specification which reads the updated sound data corresponding to the specific condition directed by acquired ID, The acoustic model construction function for the specification which builds the acoustic model depending on a specific condition with reference to the updated sound data, The acoustic model transmitting function which transmits an acoustic model via a network, By receiving an acoustic model and realizing the acoustic model update function which a verification function updates by the acoustic model which received the acoustic model referred to in the case of speech recognition, Can raise the recognition precision of speech recognition, without applying a big burden to a user, and. By receiving the acoustic model customized for specification via a network, even when a user uses two or more verification functions, a suitable acoustic model is used, and it is effective in the ability to acquire high recognition precision.

[0201]The language data acquisition function which acquires the updated language data with the recording medium which recorded the voice recognition program according to this invention, The updating language model ID acquisition function which acquires ID which specifies a language model in response to the renewal instructions of a language model, The language data read-out function for the specification which reads the updated language data corresponding to the specific condition directed by acquired ID, The language model construction function for the specification which builds the language model depending on a specific condition with reference to the updated language data, The language model transmitting function which transmits the built language model via a network, By receiving a language model and realizing the language model update function which a verification function updates by the language model which received the language model referred to in the case of speech recognition, Can raise the recognition precision of speech recognition, without applying a big burden to a user, and. By receiving the language model customized for specification via a network, even when a user uses two or more verification functions, a suitable language model is used, and it is effective in the ability to acquire high recognition precision.

[0202]The language data acquisition function which acquires the updated language data with the recording medium which recorded the voice recognition program according to this invention, The user dictionary read-out function which reads a user dictionary in response to the renewal instructions of a language model, The user dictionary dependence language model construction function to build the language model which read the updated language data and was dependent on the user dictionary, The language model transmitting function which transmits the built language model via a network, By receiving a language model and realizing the language model update function which a verification function updates by the language model which received the language model referred to in the case of speech recognition, It is effective in the ability to raise the recognition precision of speech recognition, without applying a big burden to a user, even when a user dictionary becomes large.

[0203]The language data acquisition function which acquires the updated language data with the recording medium which recorded the voice recognition program according to this invention, The user use text acquisition function which acquires the text which the user used in response to the renewal instructions of a language model, The user use text dependence language model construction function to build the language model which read the updated language data and was dependent on the text, The language model transmitting function which transmits the built language model via a network, It is effective in the ability to raise the recognition precision of speech recognition, without applying a big burden to a user by receiving a language model and realizing the language model update function which a verification function updates by the language model which received the language model referred to in the case of speech recognition.

[0204]The acoustic model ID acquisition function which acquires ID which specifies an acoustic model with the recording medium which recorded the voice recognition program according to this invention, Read acquired ID, acquire the voice data for adaptation from the inputted audio signal, and via a network, The voice acquisition function for adaptation which transmits the voice data read ID and for the acquired adaptation, The acoustic model adaptation function to carry out adaptation of the initial acoustic model before adaptation, and to match and store an adaptation finishing acoustic model in ID which transmitted using the transmitted voice data for adaptation, In response to the renewal instructions of an acoustic model, ID which the acoustic model ID acquisition function acquired via the network is received, The adaptation finishing acoustic model function preselection capability which chooses and reads the adaptation finishing acoustic model corresponding to ID which received out of the adaptation finishing acoustic model which the acoustic model adaptation function stored, The acoustic model transmitting function which transmits the read adaptation finishing acoustic model via a network, Without applying a big burden to a user by receiving an adaptation finishing acoustic model and realizing the acoustic model update function which a verification function updates by the adaptation finishing acoustic model which received the acoustic model referred to in the case of speech recognition, Can raise the recognition precision of speech recognition and by updating an acoustic model via a network by the adaptation finishing acoustic model which carried out adaptation to the user. Even when a user uses several different verification functions, a suitable acoustic model is used and it is effective in the ability to acquire high recognition precision.

[Translation done.]

DESCRIPTION OF DRAWINGS

[Brief Description of the Drawings]
[Drawing 1]It is a block diagram showing the composition of the voice recognition equipment by this embodiment of the invention 1.
[Drawing 2]It is a flow chart which shows the procedure of the update process of the acoustic model by this embodiment of the invention 1.
[Drawing 3]It is a flow chart which shows the procedure of the update process of the language model by this embodiment of the invention 1.
[Drawing 4]It is a block diagram showing the composition of the voice recognition equipment by this embodiment of the invention 2.
[Drawing 5]It is a flow chart which shows the procedure of the update process of the language model by this embodiment of the invention 2.
[Drawing 6]It is a block diagram showing the composition of the voice recognition equipment by this embodiment of the invention 3.
[Drawing 7]It is a flow chart which shows the procedure of the update process of the language model by this embodiment of the invention 3.
[Drawing 8]It is a block diagram showing the composition of the voice recognition equipment by this embodiment of the invention 4.
[Drawing 9]It is a flow chart which shows the procedure of the update process of the language model by this embodiment of the invention 4.
[Drawing 10]It is a block diagram showing the composition of the voice recognition equipment by this embodiment of the invention 5.
[Drawing 11]It is a flow chart which shows the procedure of the update process of the acoustic model by this embodiment of the invention 5.
[Drawing 12]It is a block diagram showing the composition of conventional voice recognition equipment.
[Drawing 13]It is a flow chart which shows the procedure of the conventional speech recognition processing.
[Drawing 14]It is a block diagram showing the composition of conventional voice recognition equipment.
[Drawing 15]It is a block diagram showing the composition of conventional voice recognition equipment.
[Drawing 16]It is a figure showing the example of composition of the user dictionary 601.
[Drawing 17]It is a flow chart which shows the procedure of the conventional speech recognition processing.
[Description of Notations]
10 Voice recognition equipment, 20 acoustic-model managing server, 30 language-model managing server, 100 audio signals and 101 A collation means, 102 acoustic models, 103 language models, 104 A recognition result, the renewal instructions of 105 acoustic models, a 106 sound-data acquisition means, 107 The updated sound data, a 108 acoustic-model construction means, a 109 acoustic-model storing means, 110 An acoustic model transmitting means, a 111 acoustic-model update means, the renewal instructions of 112 language models, 113 A language data acquisition means, the language data updated 114 times, a 115 language-model construction means, 116 A language model storing means, a 117 language-model transmitting means, a 118 language-model update means, The renewal language model ID acquiring means of 401, and 402 The language data reading means for specification, 403 The language model construction means for specification, 601 user dictionaries, a 602 user-dictionary reading means, 603 A user dictionary dependence language model construction means and 801 User use text acquisition means, 802 A user use text storing means and 803 A user use text dependence language model construction means, a 1001 acoustic-model ID acquiring means, and 1002 The voice acquisition means for adaptation, 1003 initial acoustic models, 1004 An acoustic model adaptation means and 1005 An adaptation finishing acoustic model storing means and 1006 Adaptation finishing acoustic model selecting means.

[Translation done.]

(19)日本国特許庁（ＪＰ）　(12)**公開特許公報（Ａ）**　(11)特許出願公開番号
特開2002－91477
（P2002－91477A）
(43)公開日　平成14年３月27日(2002.3.27)

(51)Int.Cl.⁷	識別記号	Ｆ Ｉ	テーマコード*（参考）
Ｇ１０Ｌ　15/06		Ｇ１０Ｌ　3/00	５２１Ａ　５Ｄ０１５
15/14			５２１Ｒ
15/18			５３５Ｚ
15/28			５３７Ｄ
			５７１Ａ

審査請求　未請求　請求項の数35　ＯＬ　（全 32 頁）

(21)出願番号　特願2000－280674(P2000－280674)

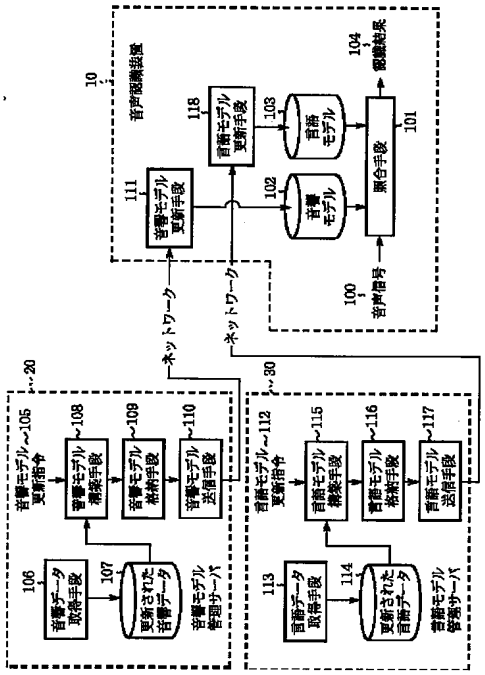
(22)出願日　平成12年９月14日(2000.9.14)

(71)出願人　000006013
三菱電機株式会社
東京都千代田区丸の内二丁目２番３号
(72)発明者　岡登　洋平
東京都千代田区丸の内二丁目２番３号　三菱電機株式会社内
(72)発明者　石井　純
東京都千代田区丸の内二丁目２番３号　三菱電機株式会社内
(74)代理人　100066474
弁理士　田澤　博昭　（外１名）

Ｆターム(参考)　5D015 GG05 HH05 HH23

(54)【発明の名称】　音声認識システム、音声認識装置、音響モデル管理サーバ、言語モデル管理サーバ、音声認識方法及び音声認識プログラムを記録したコンピュータ読み取り可能な記録媒体

(57)【要約】
【課題】　ユーザへ大きな負担をかけることなく音響モデルと言語モデルを更新し、認識精度を向上させる。
【解決手段】　音響モデル管理サーバ２０は、更新された音響データ１０７を取得し音響モデルを構築し、言語モデル管理サーバ３０は、更新された言語データ１１４を取得し言語モデルを構築して、それぞれネットワークを介して音声認識装置１０に送信する。音声認識装置１０において、音響モデル更新手段１１１は、送信された音響モデルにより音響モデル１０２を更新し、言語モデル更新手段１１８は、送信された言語モデルにより言語モデル１０３を更新する。



1

【特許請求の範囲】

【請求項１】　音声信号を入力し、音声の音響的な観測値系列の確率を求める音響モデルを参照して音声認識を行い、認識結果を出力する音声認識装置と、

上記音声認識装置とネットワークを介して接続され、更新された音響データを取得して上記音響モデルを構築する音響モデル管理サーバとを備えた音声認識システムにおいて、

上記音響モデル管理サーバが構築した上記音響モデルを上記音声認識装置に送信し、上記音声認識装置が、音声認識の際に参照する音響モデルを、上記音響モデル管理サーバが送信した音響モデルにより更新することを特徴とする音声認識システム。

【請求項２】　音響モデル管理サーバが、音声認識装置が音声認識の際に参照する音響モデルを特定するＩＤを取得し、取得したＩＤで指示される特定条件に対応して、更新された音響データを読み出し、上記特定条件に依存した音響モデルを構築して上記音声認識装置に送信することを特徴とする請求項１記載の音声認識システム。

【請求項３】　音声信号を入力し、単語列の出現確率を求める言語モデルを参照して音声認識を行い、認識結果を出力する音声認識装置と、

上記音声認識装置とネットワークを介して接続され、更新された言語データを取得して上記言語モデルを構築する言語モデル管理サーバとを備えた音声認識システムにおいて、

上記言語モデル管理サーバが構築した上記言語モデルを上記音声認識装置に送信し、上記音声認識装置が、音声認識の際に参照する言語モデルを、上記言語モデル管理サーバが送信した言語モデルにより更新することを特徴とする音声認識システム。

【請求項４】　言語モデル管理サーバが、音声認識装置が音声認識の際に参照する言語モデルを特定するＩＤを取得し、取得したＩＤで指示される特定条件に対応して、更新された言語データを読み出し、上記特定条件に依存した言語モデルを構築して上記音声認識装置に送信することを特徴とする請求項３記載の音声認識システム。

【請求項５】　音声認識装置が音声認識の際に単語を登録したユーザ辞書を参照し、

言語モデル管理サーバが、ネットワークを介して上記ユーザ辞書を読み出し、更新された言語データと、読み出した上記ユーザ辞書とを参照し、上記ユーザ辞書に依存した言語モデルを構築して上記音声認識装置に送信することを特徴とする請求項３記載の音声認識システム。

【請求項６】　言語モデル管理サーバが、音声認識装置のユーザが利用したテキストを取得し、更新された言語データと、取得した上記テキストとを参照し、上記テキストに依存した言語モデルを構築して上記音声認識装置

2

に送信することを特徴とする請求項３記載の音声認識システム。

【請求項７】　音声信号を入力し、音声の音響的な観測値系列の確率を求める音響モデルを参照して音声認識を行い、認識結果を出力する音声認識装置と、

上記音声認識装置とネットワークを介して接続され、適応化前の初期音響モデルを有する音響モデル管理サーバとを備えた音声認識システムにおいて、

上記音声認識装置が、上記音響モデルを特定するＩＤと、入力された音声信号から適応化用の音声データとを取得し、取得したＩＤ及び適応化用の音声データを、ネットワークを介して上記音響モデル管理サーバに送信し、

上記音響モデル管理サーバが、送信された適応化用の音声データを用いて、上記初期音響モデルを適応化し、適応化済み音響モデルを、送信された上記ＩＤに対応付けて格納すると共に、外部からの音響モデル更新指令を受けて、ネットワークを介して上記音声認識装置から上記音響モデルを特定するＩＤを受信し、受信したＩＤに対応する適応化済み音響モデルを、格納している適応化済み音響モデルの中から選択して読み出し、ネットワークを介して上記音声認識装置に送信し、

上記音声認識装置が、音声認識の際に参照する音響モデルを、上記音響モデル管理サーバが送信した適応化済み音響モデルにより更新することを特徴とする音声認識システム。

【請求項８】　音声の音響的な観測値系列の確率を求める音響モデルと、

音声信号を入力し上記音響モデルを参照して音声認識を行い、認識結果を出力する照合手段とを備えた音声認識装置において、

ネットワークを介して接続された音響モデル管理サーバから、更新された音響データにより構築された音響モデルを受信し、上記照合手段が音声認識の際に参照する音響モデルを、受信した音響モデルにより更新する音響モデル更新手段とを備えたことを特徴とする音声認識装置。

【請求項９】　音響モデル更新手段が、ネットワークを介して接続された音響モデル管理サーバから、更新された音響データにより構築された、照合手段が音声認識の際に参照する音響モデルの特定条件に依存した音響モデルを受信し、上記照合手段が音声認識の際に参照する音響モデルを、受信した音響モデルにより更新することを特徴とする請求項８記載の音声認識装置。

【請求項１０】　単語列の出現確率を求める言語モデルと、

音声信号を入力し上記言語モデルを参照して音声認識を行い、認識結果を出力する照合手段とを備えた音声認識装置において、

ネットワークを介して接続された言語モデル管理サーバ

から、更新された言語データにより構築された言語モデルを受信し、上記照合手段が音声認識の際に参照する言語モデルを、受信した言語モデルにより更新する言語モデル更新手段とを備えたことを特徴とする音声認識装置。

【請求項１１】　言語モデル更新手段が、ネットワークを介して接続された言語モデル管理サーバから、更新された言語データにより構築された、照合手段が音声認識の際に参照する言語モデルの特定条件に依存した言語モデルを受信し、上記照合手段が音声認識の際に参照する言語モデルを、受信した言語モデルにより更新することを特徴とする請求項１０記載の音声認識装置。

【請求項１２】　照合手段が音声認識の際に参照する単語を登録したユーザ辞書を備え、
言語モデル更新手段が、ネットワークを介して接続された言語モデル管理サーバから、更新された言語データにより構築された、上記照合手段が音声認識の際に参照するユーザ辞書に依存した言語モデルを受信し、上記照合手段が音声認識の際に参照する言語モデルを、受信した言語モデルにより更新することを特徴とする請求項１０記載の音声認識装置。

【請求項１３】　言語モデル更新手段が、ネットワークを介して接続された言語モデル管理サーバから、更新された言語データにより構築された、音声認識を行うユーザが利用したテキストに依存した言語モデルを受信し、上記照合手段が音声認識の際に参照する言語モデルを、受信した言語モデルにより更新することを特徴とする請求項１０記載の音声認識装置。

【請求項１４】　音声の音響的な観測値系列の確率を求める音響モデルと、音声信号を入力し上記音響モデルを参照して音声認識を行い、認識結果を出力する照合手段と、
上記音響モデルを特定するＩＤを取得する音響モデルＩＤ取得手段と、
上記音響モデルＩＤ取得手段が取得したＩＤを読み出し、入力された音声信号から適応化用の音声データを取得し、読み出したＩＤ及び取得した適応化用の音声データを、ネットワークを介して接続された音響モデル管理サーバに送信する適応化用音声取得手段と、
上記音響モデル管理サーバから、上記ＩＤに対応する上記適応化用の音声データにより適応化された適応化済み音響モデルを受信し、上記照合手段が音声認識の際に参照する音響モデルを、受信した適応化済み音響モデルにより更新する音響モデル更新手段とを備えたことを特徴とする音声認識装置。

【請求項１５】　更新された音響データを取得する音響データ取得手段と、
外部からの音響モデル更新指令を受けて、上記音響データ取得手段が取得した更新された音響データを読み出し、音声の音響的な観測値系列の確率を求める音響モデルを構築する音響モデル構築手段と、
上記音響モデル構築手段により構築された音響モデルを、ネットワークを介して音声認識を行う音声認識装置に送信する音響モデル送信手段とを備えたことを特徴とする音響モデル管理サーバ。

【請求項１６】　更新された音響データを取得する音響データ取得手段と、
外部からの音響モデル更新指令を受けて、ネットワークを介して接続された音声認識装置が音声認識の際に参照する音響モデルを特定するＩＤを取得する更新音響モデルＩＤ取得手段と、
上記更新音響モデルＩＤ取得手段が取得したＩＤで指示される特定条件に対応して、上記音響データ取得手段が取得した更新された音響データを読み出す特定向け音響データ読み出し手段と、
上記特定向け音響データ読み出し手段が読み出した更新された音響データを参照し、上記特定条件に依存した音響モデルを構築する特定向け音響モデル構築手段と、
上記特定向け音響モデル構築手段が構築した音響モデルを、ネットワークを介して上記音声認識装置に送信する音響モデル送信手段とを備えたことを特徴とする音響モデル管理サーバ。

【請求項１７】　音声の音響的な観測値系列の確率を求める、適応化前の初期音響モデルと、
ネットワークを介して接続された音声認識装置から送信された、適応化用の音声データと、上記音声認識装置が音声認識の際に参照する音響モデルを特定するＩＤを受信し、上記適応化用の音声データを用いて上記初期音響モデルを適応化し、適応化済み音響モデルを、受信した上記ＩＤに対応付けて適応化済み音響モデル格納手段に格納する音響モデル適応化手段と、
外部からの音響モデル更新指令を受けて、ネットワークを介して上記音声認識装置から上記ＩＤを受信し、受信したＩＤに対応する適応化済み音響モデルを、上記適応化済み音響モデル格納手段から選択して読み出す適応化済み音響モデル選択手段と、
上記適応化済み音響モデル選択手段が読み出した適応化済み音響モデルを、ネットワークを介して上記音声認識装置に送信する音響モデル送信手段とを備えたことを特徴とする音響モデル管理サーバ。

【請求項１８】　更新された言語データを取得する言語データ取得手段と、
外部からの言語モデル更新指令を受けて、上記言語データ取得手段が取得した更新された言語データを読み出し、単語列の出現確率を求める言語モデルを構築する言語モデル構築手段と、
上記言語モデル構築手段が構築した言語モデルを、ネットワークを介して音声認識を行う音声認識装置に送信する言語モデル送信手段とを備えたことを特徴とする言語モデル管理サーバ。

5

【請求項１９】　更新された言語データを取得する言語データ取得手段と、
外部からの言語モデル更新指令を受けて、ネットワークを介して接続された音声認識装置が音声認識の際に参照する言語モデルを特定するＩＤを取得する更新言語モデルＩＤ取得手段と、
上記更新言語モデルＩＤ取得手段が取得したＩＤで指示される特定条件に対応して、上記言語データ取得手段が取得した更新された言語データを読み出す特定向け言語データ読み出し手段と、
上記特定向け言語データ読み出し手段が読み出した更新された言語データを参照し、上記特定条件に依存した言語モデルを構築する特定向け言語モデル構築手段と、
上記特定向け言語モデル構築手段が構築した言語モデルを、ネットワークを介して上記音声認識装置に送信する言語モデル送信手段とを備えたことを特徴とする言語モデル管理サーバ。
【請求項２０】　更新された言語データを取得する言語データ取得手段と、
外部からの言語モデル更新指令を受けて、ネットワークを介して接続された音声認識装置が音声認識の際に参照するユーザ辞書を読み出すユーザ辞書読み出し手段と、
上記言語データ取得手段が取得した更新された言語データを読み出し、上記ユーザ辞書読み出し手段が読み出したユーザ辞書に依存した言語モデルを構築するユーザ辞書依存言語モデル構築手段と、
上記ユーザ辞書依存言語モデル構築手段が構築した言語モデルを、ネットワークを介して上記音声認識装置に送信する言語モデル送信手段とを備えたことを特徴とする言語モデル管理サーバ。
【請求項２１】　更新された言語データを取得する言語データ取得手段と、
外部からの言語モデル更新指令を受けて、ネットワークを介して接続された音声認識装置のユーザが利用したテキストを取得するユーザ利用テキスト取得手段と、
上記言語データ取得手段が取得した更新された言語データを読み出し、上記ユーザ利用テキスト取得手段が取得したテキストに依存した言語モデルを構築するユーザ利用テキスト依存言語モデル構築手段と、
上記ユーザ利用テキスト依存言語モデル構築手段が構築した言語モデルを、ネットワークを介して上記音声認識装置に送信する言語モデル送信手段とを備えたことを特徴とする言語モデル管理サーバ。
【請求項２２】　音声信号を入力し、音声の音響的な観測値系列の確率を求める音響モデルを参照して音声認識を行い、認識結果を出力する音声認識方法において、
更新された音響データを取得する第１のステップと、
音響モデル更新指令を受けて、上記第１のステップで取得した更新された音響データを読み出し、音響モデルを構築する第２のステップと、

6

上記第２のステップで構築した音響モデルを、ネットワークを介して送信する第３のステップと、
上記第３のステップで送信した音響モデルを受信し、上記音声認識の際に参照する音響モデルを、受信した音響モデルにより更新する第４のステップとを備えたことを特徴とする音声認識方法。
【請求項２３】　音声信号を入力し、単語列の出現確率を求める言語モデルを参照して音声認識を行い、認識結果を出力する音声認識方法において、
更新された言語データを取得する第１のステップと、
言語モデル更新指令を受けて、上記第１のステップで取得した更新された言語データを読み出し、言語モデルを構築する第２のステップと、
上記第２のステップで構築した言語モデルを、ネットワークを介して送信する第３のステップと、
上記第３のステップで送信した言語モデルを受信し、上記音声認識の際に参照する言語モデルを、受信した言語モデルにより更新する第４のステップとを備えたことを特徴とする音声認識方法。
【請求項２４】　音声信号を入力し、音声の音響的な観測値系列の確率を求める音響モデルを参照して音声認識を行い、認識結果を出力する音声認識方法において、
更新された音響データを取得する第１のステップと、
音響モデル更新指令を受けて、音声認識の際に参照する音響モデルを特定するＩＤを取得する第２のステップと、
上記第２のステップで取得したＩＤで指示される特定条件に対応して、上記第１のステップで取得した更新された音響データを読み出す第３のステップと、
上記第３のステップで読み出した更新された音響データを参照し、上記特定条件に依存した音響モデルを構築する第４のステップと、
上記第４のステップで構築した音響モデルを、ネットワークを介して送信する第５のステップと、
上記第５のステップで送信した音響モデルを受信し、音声認識の際に参照する音響モデルを、受信した音響モデルにより更新する第６のステップとを備えたことを特徴とする音声認識方法。
【請求項２５】　音声信号を入力し、単語列の出現確率を求める言語モデルを参照して音声認識を行い、認識結果を出力する音声認識方法において、
更新された言語データを取得する第１のステップと、
言語モデル更新指令を受けて、音声認識の際に参照する言語モデルを特定するＩＤを取得する第２のステップと、
上記第２のステップで取得したＩＤで指示される特定条件に対応して、上記第１のステップで取得した更新された言語データを読み出す第３のステップと、
上記第３のステップで読み出した更新された言語データを参照し、上記特定条件に依存した言語モデルを構築す

る第４のステップと、
上記第４のステップで構築した言語モデルを、ネットワークを介して送信する第５のステップと、
上記第５のステップで送信した言語モデルを受信し、音声認識の際に参照する言語モデルを、受信した言語モデルにより更新する第６のステップとを備えたことを特徴とする音声認識方法。
【請求項２６】　音声信号を入力し、単語列の出現確率を求める言語モデルと、単語を登録したユーザ辞書を参照して音声認識を行い、認識結果を出力する音声認識方法において、
更新された言語データを取得する第１のステップと、
言語モデル更新指令を受けて、音声認識の際に参照するユーザ辞書を読み出す第２のステップと、
上記第１のステップで取得した更新された言語データを読み出し、上記第２のステップで読み出したユーザ辞書に依存した言語モデルを構築する第３のステップと、
上記第３のステップで構築した言語モデルを、ネットワークを介して送信する第４のステップと、
上記第４のステップで送信した言語モデルを受信し、音声認識の際に参照する言語モデルを、受信した言語モデルにより更新する第５のステップとを備えたことを特徴とする音声認識方法。
【請求項２７】　音声信号を入力し、単語列の出現確率を求める言語モデルを参照して音声認識を行い、認識結果を出力する音声認識方法において、
更新された言語データを取得する第１のステップと、
言語モデル更新指令を受けて、音声認識を行うユーザが利用したテキストを取得する第２のステップと、
上記第１のステップで取得した更新された言語データを読み出し、上記第２のステップで取得したテキストに依存した言語モデルを構築する第３のステップと、
上記第３のステップで構築した言語モデルを、ネットワークを介して送信する第４のステップと、
上記第４のステップで送信した言語モデルを受信し、音声認識の際に参照する言語モデルを、受信した言語モデルにより更新する第５のステップとを備えたことを特徴とする音声認識方法。
【請求項２８】　音声信号を入力し、音声の音響的な観測値系列の確率を求める音響モデルを参照して音声認識を行い、認識結果を出力する音声認識方法において、
上記音響モデルを特定するＩＤを取得する第１のステップと、
上記第１のステップで取得したＩＤを読み出し、入力された音声信号から適応化用の音声データを取得し、ネットワークを介して、読み出したＩＤ及び取得した適応化用の音声データを送信する第２のステップと、
上記第２のステップで送信した適応化用の音声データを用いて、適応化前の初期音響モデルを適応化し、適応化済み音響モデルを、上記第２のステップで送信したＩＤに対応付けて格納する第３のステップと、
音響モデル更新指令を受けて、ネットワークを介して上記第１のステップで取得したＩＤを受信し、受信したＩＤに対応する適応化済み音響モデルを、上記第３のステップで格納している適応化済み音響モデルの中から選択して読み出す第４のステップと、
上記第４のステップで読み出した適応化済み音響モデルを、ネットワークを介して送信する第５のステップと、
上記第５のステップで送信した適応化済み音響モデルを受信し、音声認識の際に参照する音響モデルを、受信した適応化済み音響モデルにより更新する第６のステップとを備えたことを特徴とする音声認識方法。
【請求項２９】　音声信号を入力し、音声の音響的な観測値系列の確率を求める音響モデルを参照して音声認識を行い、認識結果を出力する照合機能を実現させる音声認識プログラムを記録した記録媒体であって、
更新された音響データを取得する音響データ取得機能と、
音響モデル更新指令を受けて、上記音響データ取得機能が取得した更新された音響データを読み出し、音響モデルを構築する音響モデル構築機能と、
上記音響モデル構築機能が構築した音響モデルを、ネットワークを介して送信する音響モデル送信機能と、
上記音響モデル送信機能が送信した音響モデルを受信し、上記照合機能が音声認識の際に参照する音響モデルを、受信した音響モデルにより更新する音響モデル更新機能とを実現させる音声認識プログラムを記録したコンピュータ読み取り可能な記録媒体。
【請求項３０】　音声信号を入力し、単語列の出現確率を求める言語モデルを参照して音声認識を行い、認識結果を出力する照合機能を実現させる音声認識プログラムを記録した記録媒体であって、
更新された言語データを取得する言語データ取得機能と、
言語モデル更新指令を受けて、上記言語データ取得機能が取得した更新された言語データを読み出し、言語モデルを構築する言語モデル構築機能と、
上記言語モデル構築機能が構築した言語モデルを、ネットワークを介して送信する言語モデル送信機能と、
上記言語モデル送信機能が送信した言語モデルを受信し、上記照合機能が音声認識の際に参照する言語モデルを、受信した言語モデルにより更新する言語モデル更新機能とを実現させる音声認識プログラムを記録したコンピュータ読み取り可能な記録媒体。
【請求項３１】　音声信号を入力し、音声の音響的な観測値系列の確率を求める音響モデルを参照して音声認識を行い、認識結果を出力する照合機能を実現させる音声認識プログラムを記録した記録媒体であって、
更新された音響データを取得する音響データ取得機能と、音響モデル更新指令を受けて、上記音響モデルを特

定するＩＤを取得する更新音響モデルＩＤ取得機能と、
上記更新音響モデルＩＤ取得機能が取得したＩＤで指示される特定条件に対応して、上記音響データ取得機能が取得した更新された音響データを読み出す特定向け音響データ読み出し機能と、
上記特定向け音響データ読み出し機能が読み出した更新された音響データを参照し、上記特定条件に依存した音響モデルを構築する特定向け音響モデル構築機能と、
上記特定向け音響モデル構築機能が構築した音響モデルを、ネットワークを介して送信する音響モデル送信機能と、
上記音響モデル送信機能が送信した音響モデルを受信し、上記照合機能が音声認識の際に参照する音響モデルを、受信した音響モデルにより更新する音響モデル更新機能とを実現させる音声認識プログラムを記録したコンピュータ読み取り可能な記録媒体。
【請求項３２】　音声信号を入力し、単語列の出現確率を求める言語モデルを参照して音声認識を行い、認識結果を出力する照合機能を実現させる音声認識プログラムを記録した記録媒体であって、
更新された言語データを取得する言語データ取得機能と、
言語モデル更新指令を受けて、上記言語モデルを特定するＩＤを取得する更新言語モデルＩＤ取得機能と、
上記更新言語モデルＩＤ取得機能が取得したＩＤで指示される特定条件に対応して、上記言語データ取得機能が取得した更新された言語データを読み出す特定向け言語データ読み出し機能と、
上記特定向け言語データ読み出し機能が読み出した更新された言語データを参照し、上記特定条件に依存した言語モデルを構築する特定向け言語モデル構築機能と、
上記特定向け言語モデル構築機能が構築した言語モデルを、ネットワークを介して送信する言語モデル送信機能と、
上記言語モデル送信機能が送信した言語モデルを受信し、上記照合機能が音声認識の際に参照する言語モデルを、受信した言語モデルにより更新する言語モデル更新機能とを実現させる音声認識プログラムを記録したコンピュータ読み取り可能な記録媒体。
【請求項３３】　音声信号を入力し、単語列の出現確率を求める言語モデルと、単語を登録したユーザ辞書を参照して音声認識を行い、認識結果を出力する照合機能を実現させる音声認識プログラムを記録した記録媒体であって、
更新された言語データを取得する言語データ取得機能と、
言語モデル更新指令を受けて、上記ユーザ辞書を読み出すユーザ辞書読み出し機能と、
上記言語データ取得機能が取得した更新された言語データを読み出し、上記ユーザ辞書読み出し機能が読み出したユーザ辞書に依存した言語モデルを構築するユーザ辞書依存言語モデル構築機能と、
上記ユーザ辞書依存言語モデル構築機能が構築した言語モデルを、ネットワークを介して送信する言語モデル送信機能と、
上記言語モデル送信機能が送信した言語モデルを受信し、上記照合機能が音声認識の際に参照する言語モデルを、受信した言語モデルにより更新する言語モデル更新機能とを実現させる音声認識プログラムを記録したコンピュータ読み取り可能な記録媒体。
【請求項３４】　音声信号を入力し、単語列の出現確率を求める言語モデルを参照して音声認識を行い、認識結果を出力する照合機能を実現させる音声認識プログラムを記録した記録媒体であって、
更新された言語データを取得する言語データ取得機能と、
言語モデル更新指令を受けて、音声認識を行うユーザが利用したテキストを取得するユーザ利用テキスト取得機能と、
上記言語データ取得機能が取得した更新された言語データを読み出し、上記ユーザ利用テキスト取得機能が取得したテキストに依存した言語モデルを構築するユーザ利用テキスト依存言語モデル構築機能と、
上記ユーザ利用テキスト依存言語モデル構築機能が構築した言語モデルを、ネットワークを介して送信する言語モデル送信機能と、
上記言語モデル送信機能が送信した言語モデルを受信し、上記照合機能が音声認識の際に参照する言語モデルを、受信した言語モデルにより更新する言語モデル更新機能とを実現させる音声認識プログラムを記録したコンピュータ読み取り可能な記録媒体。
【請求項３５】　音声信号を入力し、音声の音響的な観測値系列の確率を求める音響モデルを参照して音声認識を行い、認識結果を出力する照合機能を実現させる音声認識プログラムを記録した記録媒体であって、
上記音響モデルを特定するＩＤを取得する音響モデルＩＤ取得機能と、
上記音響モデルＩＤ取得機能が取得したＩＤを読み出し、入力された音声信号から適応化用の音声データを取得し、ネットワークを介して、読み出したＩＤ及び取得した適応化用の音声データを送信する適応化用音声取得機能と、
上記適応化用音声取得機能が送信した適応化用の音声データを用いて、適応化前の初期音響モデルを適応化し、適応化済み音響モデルを、上記適応化用音声取得機能が送信したＩＤに対応付けて格納する音響モデル適応化機能と、
音響モデル更新指令を受けて、ネットワークを介して上記音響モデルＩＤ取得機能が取得したＩＤを受信し、受信したＩＤに対応する適応化済み音響モデルを、上記音

響モデル適応化機能が格納した適応化済み音響モデルの中から選択して読み出す適応化済み音響モデル選択機能と、
上記適応化済み音響モデル選択機能が読み出した適応化済み音響モデルを、ネットワークを介して送信する音響モデル送信機能と、
上記音響モデル送信機能が送信した適応化済み音響モデルを受信し、上記照合機能が音声認識の際に参照する音響モデルを、受信した適応化済み音響モデルにより更新する音響モデル更新機能とを実現させる音声認識プログラムを記録したコンピュータ読み取り可能な記録媒体。

【発明の詳細な説明】

【０００１】

【発明の属する技術分野】この発明は、音声認識の際に参照する音響モデルと言語モデルを、高い認識率が得られるように、ネットワークを介して最新の状態に更新す＊る音声認識システム、音声認識装置、音響モデル管理サーバ、言語モデル管理サーバ、音声認識方法及び音声認識プログラムを記録したコンピュータ読み取り可能な記録媒体に関するものである。

【０００２】

【従来の技術】音声認識においては、通常、ディジタル化して入力された音声を、信号処理手法を用いて音声の音響的特徴を良く表すベクトルの時系列へ変換した後、音声のモデル（音響モデル、言語モデル）との照合処理を行う。

【０００３】照合処理はｎ個の時刻フレームからなる音響特徴ベクトル時系列Ａ＝［ａ１ａ２：：ａｎ］から発声された単語列Ｗ＝［ｗ１ｗ２：：ｗｋ］（ｋは単語数）を求める問題である。認識精度が最も高くなるような単語列を推定するには、出現確率Ｐ（Ｗ｜Ａ）が最大となる認識単語列Ｗ＊を求めれば良い。すなわち、

$$W^* = \arg\max_W P(W \mid A) \qquad (1)$$

※（２）式のように書き換える。

$$P(W \mid A) = P(W)\,P(A \mid W) / P(A) \qquad (2)$$

★を与えないため、右辺分子を最大化するＷを求めれば良い。すなわち、（３）式のようになる。

$$W^* = \arg\max_W P(W)\,P(A \mid W) \qquad (3)$$

☆モデルの構築では、あらかじめ多数の話者からの単語、文等の音声データを収集し、統計的手法を利用して認識精度や認識精度と良く関連した指標が向上するように推定する。例えば、音響モデルを構成する隠れマルコフモデルのパラメータを、音響モデルが学習データを出力する尤度が最大となるように、バウム・ウェルチアルゴリズムを用いて推定する。音響モデルの推定方法は、文献１下巻において詳細に記されている。

【０００８】同様に、言語モデルの構築では、新聞や会話の書き起こし等のテキストから、言語モデルの構造に従って、それぞれの発話や発話を構成する単語の出現する確率を計算する。例えば、ｎグラム言語モデルにおいて、ｎ＝２とおいたとき（バイグラム言語モデルと呼ばれる）、Ｐ（Ｗ）は（４）式のように近似される。ｎグラム言語モデルのパラメータは、学習用テキストデータ内の隣接するｎ単語の頻度から推定される。言語モデルの推定方法は、文献２において詳細に記されている。

$$\begin{aligned} &P(w1..w.k) \\ &= P(w1)\,P(w2 \mid w1)..P.(wk \mid w1..w.k-1) \\ &\fallingdotseq P(w1)\,P(w2 \mid w1)..P.(wk \mid wk-1) \end{aligned} \qquad (4)$$

ば、文字や形態素、文節等言語的な単位やエントロピー基準に基づいたテキストの分割、これらの組み合わせ等である。

【０００１】

【発明の属する技術分野】この発明は、音声認識の際に参照する音響モデルと言語モデルを、高い認識率が得られるように、ネットワークを介して最新の状態に更新す＊

$$W^* = \arg\max_W P(W \mid A)$$

ただし、Ｐ（Ｗ｜Ａ）を直接求めることは通常困難である。そこでベイズの定理を用いて、Ｐ（Ｗ｜Ａ）は　※

$$P(W \mid A) = P(W)\,P(A \mid W) / P(A)$$

【０００４】ここで、左辺を最大化するＷを求める際に、右辺分母であるＰ（Ａ）は認識候補となるＷに影響★

$$W^* = \arg\max_W P(W)\,P(A \mid W)$$

ここで、Ｐ（Ｗ）を与える確率モデルを言語モデルと呼び、Ｐ（Ａ｜Ｗ）を与える確率モデルを音響モデルと呼ぶ。

【０００５】音声認識におけるこれらの代表的なモデル化方法は、音響モデルを隠れマルコフモデルで表現し、言語モデルをｎグラムと呼ばれる単語のｎ－１重マルコフ過程で表現する方法である。

【０００６】これらの方法の詳細は、例えば「音声認識の基礎（上、下）」Ｌ．ＲＡＢＩＮＥＲ，Ｂ．Ｈ．ＪＵＡＮＧ，古井監訳、１９９５年、１１月、ＮＴＴアドバンステクノロジ（以下、文献１とする）、「確率的言語モデル」北研二、東京大学出版会（以下、文献２とする）、「音声・音情報のディジタル信号処理」鹿野清宏、中村哲、伊勢史郎共著、１９９７年、１１月、昭晃堂（以下、文献３とする）に記されている。

【０００７】これらの方法において、モデルを構成するパラメータは大量のデータから統計的に推定する。音響☆

$$\begin{aligned} &P(w1..w.k) \\ &= P(w1)\,P(w2 \mid w1)..P.(wk \mid w1..w.k-1) \\ &\fallingdotseq P(w1)\,P(w2 \mid w1)..P.(wk \mid wk-1) \end{aligned}$$

【０００９】このように大量テキストを用いて、それぞれの単語の出現確率を統計的に推定することによって、統計量を用いない方法に比べて高い認識精度を得る言語モデルを構築できる。なお、日本語ではテキストが分かち書きされないため、単語の定義はあいまいであるが、本文では整合性のある何らかの手段でテキストを分割したそれぞれの単位を単語と定義する。この単語は、例え

【００１０】図１２は上記文献１に開示された従来の音声認識装置の構成を示すブロック図である。図１２において、１０１は音声信号１００を入力し音声認識して認識結果１０４を出力する照合手段、１０２は照合手段１

０１が音声認識する際に参照する音響モデル、１０３は照合手段１０１が音声認識する際に参照する言語モデルである。

【００１１】次に動作について説明する。照合手段１０１は、ユーザの音声信号１００を入力し、音響モデル１０２及び言語モデル１０３を参照して、音声認識を実行し認識結果１０４を出力する。音響モデル１０２は、入力されたユーザの音声信号１００の音声波形を信号処理して得られる音響特徴ベクトルの時系列と、例えば、音素等で表される音声認識装置が扱う最小のシンボル情報との写像関係を表す。言語モデル１０３は、音響モデル１０２により写像されるシンボルの組み合わせにより表される単語等より長い認識単位との対応関係と、単語の出現情報を記述する。音響モデル１０２は、あるシンボルのモデルがベクトル時系列を出力する確率を求めるものであり、すなわち、音声の音響的な観測値系列の確率を求めるものであり、言語モデル１０３は、ある単語列の出現確率を求めるものである。

【００１２】図１３は音声信号１００を入力して認識結果１０４を得る従来の音声認識処理の手順を示すフローチャートである。ステップＳＴ１３０１において、入力された音声信号１００はＡ／Ｄ変換されてディジタル信号となる。ステップＳＴ１３０２において、ディジタル化された音声信号は適当な間隔をおいて信号処理され、音声の性質をよく表す音響特徴ベクトルの時系列へと変換される。

【００１３】ステップＳＴ１３０３において、音響特徴ベクトルは音響照合処理により音響モデル１０２と照合され、それぞれの認識候補について、音響特徴ベクトルの時系列を出力する確率が求められる。ステップＳＴ１３０４において、それぞれの認識候補はさらに言語照合処理によって、言語モデル１０３と照合され単語列の出力確率が乗じられる。最後にステップＳＴ１３０５において、それぞれの認識候補から最も適切な候補を選択して認識結果１０４を得る。通常、最も適切な認識結果とは、上記照合によって最も確率が高いとされた認識候補である。

【００１４】上記方法により構成された音響モデル及び言語モデルのみでは、十分な性能が達成されない場合で、ユーザがカスタマイズ可能な音声認識装置では、認識されにくい音声や単語をより良く認識させるために、音響モデルをユーザに適応化させたり、認識対象単語をユーザ辞書に追加することによって、認識精度を高めることができる場合がある。

【００１５】まず、音響モデルを適応化させる場合について説明する。図１４は上記文献１に開示された、音響モデルを音声信号１００に適応化させる音響モデル適応化手段を備えた従来の音声認識装置の構成を示すブロック図である。図１４において、図１２と異なる部分は、入力する音声信号１００に対して音響モデルを適応させるために、初期音響モデル１００３，音響モデル適応化手段１００４，適応化済み音響モデル１４０１を備えることと、音声信号１００が音響モデル適応化手段１００４及び照合手段１０１により選択されることである。

【００１６】次に図１４に示す音声認識装置の動作について説明する。音響モデル適応化手段１００４は、実際の認識前に収集した適応化用音声（音声信号１００）と初期音響モデル１００３から、例えば最大事後確率推定法を用いて、初期音響モデル１００３を適応化用音声に適応化させ、適応化済み音響モデル１４０１を得る。音響モデルの適応化方法については、文献３の７章に示されている。照合手段１０１は、音声信号１００を入力し、適応化済み音響モデル１４０１及び言語モデル１０３を参照して、音声認識を行い認識結果１０４を出力する。

【００１７】次に認識対象単語をユーザが辞書に登録する場合について説明する。図１５は上記文献１に開示された、ユーザ辞書が追加された従来の音声認識装置の構成を示すブロック図である。図１５において、図１２と異なる部分はユーザ辞書６０１を備えることである。

【００１８】図１６はユーザ辞書６０１の構成例を示す図である。ユーザ辞書６０１は認識されない単語、認識されにくい単語を、より良く認識させるためにユーザが登録した単語の集まりで、単語の表記、読みからなる単語の一覧である。

【００１９】次に図１５に示す音声認識装置の動作について説明する。図１７はユーザ辞書６０１が追加された従来の音声認識処理の手順を示すフローチャートである。図１７において、図１３と異なる点は、ステップＳＴ１７０４の言語照合処理において、言語モデル１０３とユーザ辞書６０１を参照するために、ユーザ辞書６０１に登録された単語が認識対象単語に加えられることである。

【００２０】ユーザ辞書６０１に登録された単語は、適当な接続確率によって、任意の単語列と接続可能とする。例えば、単語の出現条件を先行する１単語のみにより決定するバイグラム言語モデルにおいて、任意の単語ｗｉとｗｊにはさまれるユーザ辞書登録単語ｗｕｓｅｒの確率Ｐ（ｗｕｓｅｒ｜ｗｉ），Ｐ（ｗｊ｜ｗｕｓｅｒ）に一定値を与え、言語モデル全体の確率が１になるように確率値を再配分する。この結果、ユーザは登録した単語が含まれた認識結果を得られる。

【００２１】しかし、図１４及び図１５に示す音声認識装置の構成では、認識精度を高めるために、音声信号１００と適応化済み音響モデル１４０１を作成したり、ユーザ辞書６０１に単語を登録することによって、ユーザが音響モデル１０２，言語モデル１０３を自らカスタマイズする必要があり、ユーザに大きな負担を強いることになる。

【００２２】また、ユーザ辞書６０１へ単語を登録した

場合であっても、言語モデル構築時点では、出現していなかった単語や考慮されていなかった用法があるために、それらの単語へ付与した出現確率が不適切である場合がある。さらに、これが原因となり認識精度が低下する可能性がある。

【００２３】さらに、上記のようにカスタマイズされた音響モデル１０２及び言語モデル１０３は、特定の照合手段１０１からのみ参照される。このため、それ以外の照合手段１０１が、このカスタマイズした音響モデル１０２及び言語モデル１０３を利用すると、認識精度が低下してしまう。

【００２４】このため、漢字かな変換や機械翻訳等の言語処理システムでは、上記に示した問題のうち、ユーザ辞書６０１の更新に関し、ネットワークを介して辞書を自動的に更新する機能を備えることによって、ユーザの負担を軽減したシステムが、例えば特開平１０－２６０９６０号公報のように提案されている。しかし、上記公報では、音声のようなパターン認識のためのモデルを扱うことは考慮しておらず、音響モデル１０２のようなパターン情報に関するモデルに適用することができない。

【００２５】さらに、言語モデル１０３の更新においても、ユーザの登録単語数が増加すると、適切でない短い単語が挿入された湧き出し誤りを生じやすくなることから、認識精度の低下が起きやすくなる。この認識精度の低下は、ユーザ辞書６０１へ追加登録した単語が、言語モデル１０３を構築した時点では存在しなかったり、使用環境が変化しているために、例えばユーザ登録単語の出現確率Ｐ（ｗｕｓｅｒ｜ｗｉ），Ｐ（ｗｊ｜ｗｕｓｅｒ）等に適切でない出現確率が付与されている場合があるからである。この結果、不適切な認識結果を得やすくなり、認識精度が低下する可能性がある。これを防ぐためには、単語の出現確率を適切に設定する必要があるが、ユーザ自身が妥当な値を与えることは一般に困難である。

【００２６】

【発明が解決しようとする課題】従来の音声認識装置は以上のように構成されているので、認識精度を高めるために音響モデルや言語モデルをカスタマイズする際に、ユーザに大きな負担をかけるという課題があった。

【００２７】また、ユーザごとにカスタマイズされた音声認識装置以外を利用する場合、認識精度が低下するという課題があった。

【００２８】さらに、ネットワークを介したカスタマイズによって、音響モデルを自動更新することができないという課題があった。

【００２９】さらに、ユーザ辞書への登録が増加すると認識精度が低下しやすいという課題があった。

【００３０】この発明は上記のような課題を解決するためになされたもので、ネットワークに接続されたサーバ側で、最新の音響データ又は言語データを取得して、最新の状態にある音響モデル又は言語モデルを構築したり、ユーザに対応した音響モデル又は言語モデルを構築し、ネットワークを介してユーザ側の音響モデルと言語モデルを更新することで、ユーザに大きな負担をかけることなく認識精度を向上させる音声認識システム、音声認識装置、音響モデル管理サーバ、言語モデル管理サーバ、音声認識方法及び音声認識プログラムを記録したコンピュータ読み取り可能な記録媒体を得ることを目的とする。

【００３１】また、ネットワークを介して接続することによって、あらゆる音声認識装置において、カスタマイズされた音響モデル又は言語モデルを利用できる音声認識システム、音声認識装置、音響モデル管理サーバ、言語モデル管理サーバ、音声認識方法及び音声認識プログラムを記録したコンピュータ読み取り可能な記録媒体を得ることを目的とする。

【００３２】さらに、ユーザから得られる辞書やテキストと半自動的に収集されたテキストを利用することによって、ユーザ辞書が大きくなった場合でも、認識精度が低下しにくい音声認識システム、音声認識装置、音響モデル管理サーバ、言語モデル管理サーバ、音声認識方法及び音声認識プログラムを記録したコンピュータ読み取り可能な記録媒体を得ることを目的とする。

【００３３】

【課題を解決するための手段】この発明に係る音声認識システムは、音声信号を入力し、音声の音響的な観測値系列の確率を求める音響モデルを参照して音声認識を行い、認識結果を出力する音声認識装置と、上記音声認識装置とネットワークを介して接続され、更新された音響データを取得して上記音響モデルを構築する音響モデル管理サーバとを備えたものにおいて、上記音響モデル管理サーバが構築した上記音響モデルを上記音声認識装置に送信し、上記音声認識装置が、音声認識の際に参照する音響モデルを、上記音響モデル管理サーバが送信した音響モデルにより更新するものである。

【００３４】この発明に係る音声認識システムは、音響モデル管理サーバが、音声認識装置が音声認識の際に参照する音響モデルを特定するＩＤを取得し、取得したＩＤで指示される特定条件に対応して、更新された音響データを読み出し、上記特定条件に依存した音響モデルを構築して上記音声認識装置に送信するものである。

【００３５】この発明に係る音声認識システムは、音声信号を入力し、単語列の出現確率を求める言語モデルを参照して音声認識を行い、認識結果を出力する音声認識装置と、上記音声認識装置とネットワークを介して接続され、更新された言語データを取得して上記言語モデルを構築する言語モデル管理サーバとを備えたものにおいて、上記言語モデル管理サーバが構築した上記言語モデルを上記音声認識装置に送信し、上記音声認識装置が、音声認識の際に参照する言語モデルを、上記言語モデル

管理サーバが送信した言語モデルにより更新するものである。

【００３６】この発明に係る音声認識システムは、言語モデル管理サーバが、音声認識装置が音声認識の際に参照する言語モデルを特定するＩＤを取得し、取得したＩＤで指示される特定条件に対応して、更新された言語データを読み出し、上記特定条件に依存した言語モデルを構築して上記音声認識装置に送信するものである。

【００３７】この発明に係る音声認識システムは、音声認識装置が音声認識の際に単語を登録したユーザ辞書を参照し、言語モデル管理サーバが、ネットワークを介して上記ユーザ辞書を読み出し、更新された言語データと、読み出した上記ユーザ辞書とを参照し、上記ユーザ辞書に依存した言語モデルを構築して上記音声認識装置に送信するものである。

【００３８】この発明に係る音声認識システムは、言語モデル管理サーバが、音声認識装置のユーザが利用したテキストを取得し、更新された言語データと、取得した上記テキストとを参照し、上記テキストに依存した言語モデルを構築して上記音声認識装置に送信するものである。

【００３９】この発明に係る音声認識システムは、音声信号を入力し、音声の音響的な観測値系列の確率を求める音響モデルを参照して音声認識を行い、認識結果を出力する音声認識装置と、上記音声認識装置とネットワークを介して接続され、適応化前の初期音響モデルを有する音響モデル管理サーバとを備えたものにおいて、上記音声認識装置が、上記音響モデルを特定するＩＤと、入力された音声信号から適応化用の音声データとを取得し、取得したＩＤ及び適応化用の音声データを、ネットワークを介して上記音響モデル管理サーバに送信し、上記音響モデル管理サーバが、送信された適応化用の音声データを用いて、上記初期音響モデルを適応化し、適応化済み音響モデルを、送信された上記ＩＤに対応付けて格納すると共に、外部からの音響モデル更新指令を受けて、ネットワークを介して上記音声認識装置から上記音響モデルを特定するＩＤを受信し、受信したＩＤに対応する適応化済み音響モデルを、格納している適応化済み音響モデルの中から選択して読み出し、ネットワークを介して上記音声認識装置に送信し、上記音声認識装置が、音声認識の際に参照する音響モデルを、上記音響モデル管理サーバが送信した適応化済み音響モデルにより更新するものである。

【００４０】この発明に係る音声認識装置は、音声の音響的な観測値系列の確率を求める音響モデルと、音声信号を入力し上記音響モデルを参照して音声認識を行い、認識結果を出力する照合手段とを備えたものにおいて、ネットワークを介して接続された音響モデル管理サーバから、更新された音響データにより構築された音響モデルを受信し、上記照合手段が音声認識の際に参照する音響モデルを、受信した音響モデルにより更新する音響モデル更新手段とを備えたものである。

【００４１】この発明に係る音声認識装置は、音響モデル更新手段が、ネットワークを介して接続された音響モデル管理サーバから、更新された音響データにより構築された、照合手段が音声認識の際に参照する音響モデルの特定条件に依存した音響モデルを受信し、上記照合手段が音声認識の際に参照する音響モデルを、受信した音響モデルにより更新するものである。

【００４２】この発明に係る音声認識装置は、単語列の出現確率を求める言語モデルと、音声信号を入力し上記言語モデルを参照して音声認識を行い、認識結果を出力する照合手段とを備えたものにおいて、ネットワークを介して接続された言語モデル管理サーバから、更新された言語データにより構築された言語モデルを受信し、上記照合手段が音声認識の際に参照する言語モデルを、受信した言語モデルにより更新する言語モデル更新手段とを備えたものである。

【００４３】この発明に係る音声認識装置は、言語モデル更新手段が、ネットワークを介して接続された言語モデル管理サーバから、更新された言語データにより構築された、照合手段が音声認識の際に参照する言語モデルの特定条件に依存した言語モデルを受信し、上記照合手段が音声認識の際に参照する言語モデルを、受信した言語モデルにより更新するものである。

【００４４】この発明に係る音声認識装置は、照合手段が音声認識の際に参照する単語を登録したユーザ辞書を備え、言語モデル更新手段が、ネットワークを介して接続された言語モデル管理サーバから、更新された言語データにより構築された、上記照合手段が音声認識の際に参照するユーザ辞書に依存した言語モデルを受信し、上記照合手段が音声認識の際に参照する言語モデルを、受信した言語モデルにより更新するものである。

【００４５】この発明に係る音声認識装置は、言語モデル更新手段が、ネットワークを介して接続された言語モデル管理サーバから、更新された言語データにより構築された、音声認識を行うユーザが利用したテキストに依存した言語モデルを受信し、上記照合手段が音声認識の際に参照する言語モデルを、受信した言語モデルにより更新するものである。

【００４６】この発明に係る音声認識装置は、音声の音響的な観測値系列の確率を求める音響モデルと、音声信号を入力し上記音響モデルを参照して音声認識を行い、認識結果を出力する照合手段と、上記音響モデルを特定するＩＤを取得する音響モデルＩＤ取得手段と、上記音響モデルＩＤ取得手段が取得したＩＤを読み出し、入力された音声信号から適応化用の音声データを取得し、読み出したＩＤ及び取得した適応化用の音声データを、ネットワークを介して接続された音響モデル管理サーバに送信する適応化用音声取得手段と、上記音響モデル管理

サーバから、上記ＩＤに対応する上記適応化用の音声データにより適応化された適応化済み音響モデルを受信し、上記照合手段が音声認識の際に参照する音響モデルを、受信した適応化済み音響モデルにより更新する音響モデル更新手段とを備えたものである。

【００４７】この発明に係る音響モデル管理サーバは、更新された音響データを取得する音響データ取得手段と、外部からの音響モデル更新指令を受けて、上記音響データ取得手段が取得した更新された音響データを読み出し、音声の音響的な観測値系列の確率を求める音響モデルを構築する音響モデル構築手段と、上記音響モデル構築手段により構築された音響モデルを、ネットワークを介して音声認識を行う音声認識装置に送信する音響モデル送信手段とを備えたものである。

【００４８】この発明に係る音響モデル管理サーバは、更新された音響データを取得する音響データ取得手段と、外部からの音響モデル更新指令を受けて、ネットワークを介して接続された音声認識装置が音声認識の際に参照する音響モデルを特定するＩＤを取得する更新音響モデルＩＤ取得手段と、上記更新音響モデルＩＤ取得手段が取得したＩＤで指示される特定条件に対応して、上記音響データ取得手段が取得した更新された音響データを読み出す特定向け音響データ読み出し手段と、上記特定向け音響データ読み出し手段が読み出した更新された音響データを参照し、上記特定条件に依存した音響モデルを構築する特定向け音響モデル構築手段と、上記特定向け音響モデル構築手段が構築した音響モデルを、ネットワークを介して上記音声認識装置に送信する音響モデル送信手段とを備えたものである。

【００４９】この発明に係る音響モデル管理サーバは、音声の音響的な観測値系列の確率を求める、適応化前の初期音響モデルと、ネットワークを介して接続された音声認識装置から送信された、適応化用の音声データと、上記音声認識装置が音声認識の際に参照する音響モデルを特定するＩＤを受信し、上記適応化用の音声データを用いて上記初期音響モデルを適応化し、適応化済み音響モデルを、受信した上記ＩＤに対応付けて適応化済み音響モデル格納手段に格納する音響モデル適応化手段と、外部からの音響モデル更新指令を受けて、ネットワークを介して上記音声認識装置から上記ＩＤを受信し、受信したＩＤに対応する適応化済み音響モデルを、上記適応化済み音響モデル格納手段から選択して読み出す適応化済み音響モデル選択手段と、上記適応化済み音響モデル選択手段が読み出した適応化済み音響モデルを、ネットワークを介して上記音声認識装置に送信する音響モデル送信手段とを備えたものである。

【００５０】この発明に係る言語モデル管理サーバは、更新された言語データを取得する言語データ取得手段と、外部からの言語モデル更新指令を受けて、上記言語データ取得手段が取得した更新された言語データを読み出し、単語列の出現確率を求める言語モデルを構築する言語モデル構築手段と、上記言語モデル構築手段が構築した言語モデルを、ネットワークを介して音声認識を行う音声認識装置に送信する言語モデル送信手段とを備えたものである。

【００５１】この発明に係る言語モデル管理サーバは、更新された言語データを取得する言語データ取得手段と、外部からの言語モデル更新指令を受けて、ネットワークを介して接続された音声認識装置が音声認識の際に参照する言語モデルを特定するＩＤを取得する更新言語モデルＩＤ取得手段と、上記更新言語モデルＩＤ取得手段が取得したＩＤで指示される特定条件に対応して、上記言語データ取得手段が取得した更新された言語データを読み出す特定向け言語データ読み出し手段と、上記特定向け言語データ読み出し手段が読み出した更新された言語データを参照し、上記特定条件に依存した言語モデルを構築する特定向け言語モデル構築手段と、上記特定向け言語モデル構築手段が構築した言語モデルを、ネットワークを介して上記音声認識装置に送信する言語モデル送信手段とを備えたものである。

【００５２】この発明に係る言語モデル管理サーバは、更新された言語データを取得する言語データ取得手段と、外部からの言語モデル更新指令を受けて、ネットワークを介して接続された音声認識装置が音声認識の際に参照するユーザ辞書を読み出すユーザ辞書読み出し手段と、上記言語データ取得手段が取得した更新された言語データを読み出し、上記ユーザ辞書読み出し手段が読み出したユーザ辞書に依存した言語モデルを構築するユーザ辞書依存言語モデル構築手段と、上記ユーザ辞書依存言語モデル構築手段が構築した言語モデルを、ネットワークを介して上記音声認識装置に送信する言語モデル送信手段とを備えたものである。

【００５３】この発明に係る言語モデル管理サーバは、更新された言語データを取得する言語データ取得手段と、外部からの言語モデル更新指令を受けて、ネットワークを介して接続された音声認識装置のユーザが利用したテキストを取得するユーザ利用テキスト取得手段と、上記言語データ取得手段が取得した更新された言語データを読み出し、上記ユーザ利用テキスト取得手段が取得したテキストに依存した言語モデルを構築するユーザ利用テキスト依存言語モデル構築手段と、上記ユーザ利用テキスト依存言語モデル構築手段が構築した言語モデルを、ネットワークを介して上記音声認識装置に送信する言語モデル送信手段とを備えたものである。

【００５４】この発明に係る音声認識方法は、音声信号を入力し、音声の音響的な観測値系列の確率を求める音響モデルを参照して音声認識を行い、認識結果を出力するものにおいて、更新された音響データを取得する第１のステップと、音響モデル更新指令を受けて、上記第１のステップで取得した更新された音響データを読み出

し、音響モデルを構築する第２のステップと、上記第２のステップで構築した音響モデルを、ネットワークを介して送信する第３のステップと、上記第３のステップで送信した音響モデルを受信し、上記音声認識の際に参照する音響モデルを、受信した音響モデルにより更新する第４のステップとを備えたものである。

【００５５】この発明に係る音声認識方法は、音声信号を入力し、単語列の出現確率を求める言語モデルを参照して音声認識を行い、認識結果を出力するものにおいて、更新された言語データを取得する第１のステップと、言語モデル更新指令を受けて、上記第１のステップで取得した更新された言語データを読み出し、言語モデルを構築する第２のステップと、上記第２のステップで構築した言語モデルを、ネットワークを介して送信する第３のステップと、上記第３のステップで送信した言語モデルを受信し、上記音声認識の際に参照する言語モデルを、受信した言語モデルにより更新する第４のステップとを備えたものである。

【００５６】この発明に係る音声認識方法は、音声信号を入力し、音声の音響的な観測値系列の確率を求める音響モデルを参照して音声認識を行い、認識結果を出力するものにおいて、更新された音響データを取得する第１のステップと、音響モデル更新指令を受けて、音声認識の際に参照する音響モデルを特定するＩＤを取得する第２のステップと、上記第２のステップで取得したＩＤで指示される特定条件に対応して、上記第１のステップで取得した更新された音響データを読み出す第３のステップと、上記第３のステップで読み出した更新された音響データを参照し、上記特定条件に依存した音響モデルを構築する第４のステップと、上記第４のステップで構築した音響モデルを、ネットワークを介して送信する第５のステップと、上記第５のステップで送信した音響モデルを受信し、音声認識の際に参照する音響モデルを、受信した音響モデルにより更新する第６のステップとを備えたものである。

【００５７】この発明に係る音声認識方法は、音声信号を入力し、単語列の出現確率を求める言語モデルを参照して音声認識を行い、認識結果を出力するものにおいて、更新された言語データを取得する第１のステップと、言語モデル更新指令を受けて、音声認識の際に参照する言語モデルを特定するＩＤを取得する第２のステップと、上記第２のステップで取得したＩＤで指示される特定条件に対応して、上記第１のステップで取得した更新された言語データを読み出す第３のステップと、上記第３のステップで読み出した更新された言語データを参照し、上記特定条件に依存した言語モデルを構築する第４のステップと、上記第４のステップで構築した言語モデルを、ネットワークを介して送信する第５のステップと、上記第５のステップで送信した言語モデルを受信し、音声認識の際に参照する言語モデルを、受信した言語モデルにより更新する第６のステップとを備えたものである。

【００５８】この発明に係る音声認識方法は、音声信号を入力し、単語列の出現確率を求める言語モデルと、単語を登録したユーザ辞書を参照して音声認識を行い、認識結果を出力するものにおいて、更新された言語データを取得する第１のステップと、言語モデル更新指令を受けて、音声認識の際に参照するユーザ辞書を読み出す第２のステップと、上記第１のステップで取得した更新された言語データを読み出し、上記第２のステップで読み出したユーザ辞書に依存した言語モデルを構築する第３のステップと、上記第３のステップで構築した言語モデルを、ネットワークを介して送信する第４のステップと、上記第４のステップで送信した言語モデルを受信し、音声認識の際に参照する言語モデルを、受信した言語モデルにより更新する第５のステップとを備えたものである。

【００５９】この発明に係る音声認識方法は、音声信号を入力し、単語列の出現確率を求める言語モデルを参照して音声認識を行い、認識結果を出力するものにおいて、更新された言語データを取得する第１のステップと、言語モデル更新指令を受けて、音声認識を行うユーザが利用したテキストを取得する第２のステップと、上記第１のステップで取得した更新された言語データを読み出し、上記第２のステップで取得したテキストに依存した言語モデルを構築する第３のステップと、上記第３のステップで構築した言語モデルを、ネットワークを介して送信する第４のステップと、上記第４のステップで送信した言語モデルを受信し、音声認識の際に参照する言語モデルを、受信した言語モデルにより更新する第５のステップとを備えたものである。

【００６０】この発明に係る音声認識方法は、音声信号を入力し、音声の音響的な観測値系列の確率を求める音響モデルを参照して音声認識を行い、認識結果を出力するものにおいて、上記音響モデルを特定するＩＤを取得する第１のステップと、上記第１のステップで取得したＩＤを読み出し、入力された音声信号から適応化用の音声データを取得し、ネットワークを介して、読み出したＩＤ及び取得した適応化用の音声データを送信する第２のステップと、上記第２のステップで送信した適応化用の音声データを用いて、適応化前の初期音響モデルを適応化し、適応化済み音響モデルを、上記第２のステップで送信したＩＤに対応付けて格納する第３のステップと、音響モデル更新指令を受けて、ネットワークを介して上記第１のステップで取得したＩＤを受信し、受信したＩＤに対応する適応化済み音響モデルを、上記第３のステップで格納している適応化済み音響モデルの中から選択して読み出す第４のステップと、上記第４のステップで読み出した適応化済み音響モデルを、ネットワークを介して送信する第５のステップと、上記第５のステッ

ブで送信した適応化済み音響モデルを受信し、音声認識の際に参照する音響モデルを、受信した適応化済み音響モデルにより更新する第６のステップとを備えたものである。

【００６１】この発明に係る音声認識プログラムを記録したコンピュータ読み取り可能な記録媒体は、音声信号を入力し、音声の音響的な観測値系列の確率を求める音響モデルを参照して音声認識を行い、認識結果を出力する照合機能を実現させるものであって、更新された音響データを取得する音響データ取得機能と、音響モデル更新指令を受けて、上記音響データ取得機能が取得した更新された音響データを読み出し、音響モデルを構築する音響モデル構築機能と、上記音響モデル構築機能が構築した音響モデルを、ネットワークを介して送信する音響モデル送信機能と、上記音響モデル送信機能が送信した音響モデルを受信し、上記照合機能が音声認識の際に参照する音響モデルを、受信した音響モデルにより更新する音響モデル更新機能とを実現させるものである。

【００６２】この発明に係る音声認識プログラムを記録したコンピュータ読み取り可能な記録媒体は、音声信号を入力し、単語列の出現確率を求める言語モデルを参照して音声認識を行い、認識結果を出力する照合機能を実現させるものであって、更新された言語データを取得する言語データ取得機能と、言語モデル更新指令を受けて、上記言語データ取得機能が取得した更新された言語データを読み出し、言語モデルを構築する言語モデル構築機能と、上記言語モデル構築機能が構築した言語モデルを、ネットワークを介して送信する言語モデル送信機能と、上記言語モデル送信機能が送信した言語モデルを受信し、上記照合機能が音声認識の際に参照する言語モデルを、受信した言語モデルにより更新する言語モデル更新機能とを実現させるものである。

【００６３】この発明に係る音声認識プログラムを記録したコンピュータ読み取り可能な記録媒体は、音声信号を入力し、音声の音響的な観測値系列の確率を求める音響モデルを参照して音声認識を行い、認識結果を出力する照合機能を実現させるものであって、更新された音響データを取得する音響データ取得機能と、音響モデル更新指令を受けて、上記音響モデルを特定するＩＤを取得する更新音響モデルＩＤ取得機能と、上記更新音響モデルＩＤ取得機能が取得したＩＤで指示される特定条件に対応して、上記音響データ取得機能が取得した更新された音響データを読み出す特定向け音響データ読み出し機能と、上記特定向け音響データ読み出し機能が読み出した更新された音響データを参照し、上記特定条件に依存した音響モデルを構築する特定向け音響モデル構築機能と、上記特定向け音響モデル構築機能が構築した音響モデルを、ネットワークを介して送信する音響モデル送信機能と、上記音響モデル送信機能が送信した音響モデルを受信し、上記照合機能が音声認識の際に参照する音響モデルを、受信した音響モデルにより更新する音響モデル更新機能とを実現させるものである。

【００６４】この発明に係る音声認識プログラムを記録したコンピュータ読み取り可能な記録媒体は、音声信号を入力し、単語列の出現確率を求める言語モデルを参照して音声認識を行い、認識結果を出力する照合機能を実現させるものであって、更新された言語データを取得する言語データ取得機能と、言語モデル更新指令を受けて、上記言語モデルを特定するＩＤを取得する更新言語モデルＩＤ取得機能と、上記更新言語モデルＩＤ取得機能が取得したＩＤで指示される特定条件に対応して、上記言語データ取得機能が取得した更新された言語データを読み出す特定向け言語データ読み出し機能と、上記特定向け言語データ読み出し機能が読み出した更新された言語データを参照し、上記特定条件に依存した言語モデルを構築する特定向け言語モデル構築機能と、上記特定向け言語モデル構築機能が構築した言語モデルを、ネットワークを介して送信する言語モデル送信機能と、上記言語モデル送信機能が送信した言語モデルを受信し、上記照合機能が音声認識の際に参照する言語モデルを、受信した言語モデルにより更新する言語モデル更新機能とを実現させるものである。

【００６５】この発明に係る音声認識プログラムを記録したコンピュータ読み取り可能な記録媒体は、音声信号を入力し、単語列の出現確率を求める言語モデルと、単語を登録したユーザ辞書を参照して音声認識を行い、認識結果を出力する照合機能を実現させるものであって、更新された言語データを取得する言語データ取得機能と、言語モデル更新指令を受けて、上記ユーザ辞書を読み出すユーザ辞書読み出し機能と、上記言語データ取得機能が取得した更新された言語データを読み出し、上記ユーザ辞書読み出し機能が読み出したユーザ辞書に依存した言語モデルを構築するユーザ辞書依存言語モデル構築機能と、上記ユーザ辞書依存言語モデル構築機能が構築した言語モデルを、ネットワークを介して送信する言語モデル送信機能と、上記言語モデル送信機能が送信した言語モデルを受信し、上記照合機能が音声認識の際に参照する言語モデルを、受信した言語モデルにより更新する言語モデル更新機能とを実現させるものである。

【００６６】この発明に係る音声認識プログラムを記録したコンピュータ読み取り可能な記録媒体は、音声信号を入力し、単語列の出現確率を求める言語モデルを参照して音声認識を行い、認識結果を出力する照合機能を実現させるものであって、更新された言語データを取得する言語データ取得機能と、言語モデル更新指令を受けて、音声認識を行うユーザが利用したテキストを取得するユーザ利用テキスト取得機能と、上記言語データ取得機能が取得した更新された言語データを読み出し、上記ユーザ利用テキスト取得機能が取得したテキストに依存した言語モデルを構築するユーザ利用テキスト依存言語

モデル構築機能と、上記ユーザ利用テキスト依存言語モデル構築機能が構築した言語モデルを、ネットワークを介して送信する言語モデル送信機能と、上記言語モデル送信機能が送信した言語モデルを受信し、上記照合機能が音声認識の際に参照する言語モデルを、受信した言語モデルにより更新する言語モデル更新機能とを実現させるものである。

【００６７】この発明に係る音声認識プログラムを記録したコンピュータ読み取り可能な記録媒体は、音声信号を入力し、音声の音響的な観測値系列の確率を求める音響モデルを参照して音声認識を行い、認識結果を出力する照合機能を実現させるものであって、上記音響モデルを特定するＩＤを取得する音響モデルＩＤ取得機能と、上記音響モデルＩＤ取得機能が取得したＩＤを読み出し、入力された音声信号から適応化用の音声データを取得し、ネットワークを介して、読み出したＩＤ及び取得した適応化用の音声データを送信する適応化用音声取得機能と、上記適応化用音声取得機能が送信した適応化用の音声データを用いて、適応化前の初期音響モデルを適応化し、適応化済み音響モデルを、上記適応化用音声取得機能が送信したＩＤに対応付けて格納する音響モデル適応化機能と、音響モデル更新指令を受けて、ネットワークを介して上記音響モデルＩＤ取得機能が取得したＩＤを受信し、受信したＩＤに対応する適応化済み音響モデルを、上記音響モデル適応化機能が格納した適応化済み音響モデルの中から選択して読み出す適応化済み音響モデル選択機能と、上記適応化済み音響モデル選択機能が読み出した適応化済み音響モデルを、ネットワークを介して送信する音響モデル送信機能と、上記音響モデル送信機能が送信した適応化済み音響モデルを受信し、上記照合機能が音声認識の際に参照する音響モデルを、受信した適応化済み音響モデルにより更新する音響モデル更新機能とを実現させるものである。

【００６８】

【発明の実施の形態】以下、この発明の実施の一形態を説明する。

実施の形態１．図１はこの発明の実施の形態１による音声認識システムの構成を示すブロック図である。図において、１０は音声認識を行う音声認識装置、２０はネットワークに接続されている音響モデル管理サーバ、３０はネットワークに接続されている言語モデル管理サーバである。ここで、ネットワークとは、有線あるいは無線によってディジタル信号を伝達可能な通信経路一般を示す。

【００６９】音声認識装置１０において、１００は入力する音声信号、１０１は音声信号１００の音声認識を行う照合手段、１０２は照合手段１０１が音声認識の際に参照する音響モデル、１０３は照合手段１０１が音声認識の際に参照する言語モデル、１０４は照合手段１０１が出力する認識結果、１１１はネットワークを介して送信された音響モデルにより音響モデル１０２を更新する音響モデル更新手段、１１８はネットワークを介して送信された言語モデルにより言語モデル１０３を更新する言語モデル更新手段である。

【００７０】音響モデル管理サーバ２０において、１０５は外部より与えられる音響モデル更新指令、１０６は更新された音響データを取得する音響データ取得手段、１０７は音響データ取得手段１０６が取得した更新された音響データ、１０８は音響モデル更新指令１０５を受けて、更新された音響データ１０７を読み出し、統計的手法を用いてパラメータ推定を行い音響モデルを構築する音響モデル構築手段、１０９は構築された音響モデルを格納する音響モデル格納手段、１１０は音響モデル格納手段１０９に格納されている音響モデルを、ネットワークを介して音声認識装置１０に送信する音響モデル送信手段である。

【００７１】言語モデル管理サーバ３０において、１１２は外部より与えられる言語モデル更新指令、１１３は更新された言語データを取得する言語データ取得手段、１１４は言語データ取得手段１１３が取得した更新された言語データ、１１５は言語モデル更新指令１１２を受けて、更新された言語データ１１４を読み出し、統計的手法を用いてパラメータ推定を行い言語モデルを構築する言語モデル構築手段、１１６は構築された言語モデルを格納する言語モデル格納手段、１１７は言語モデル格納手段１１６に格納されている言語モデルを、ネットワークを介して音声認識装置１０に送信する言語モデル送信手段である。

【００７２】従来技術と異なるこの発明の特徴的な部分は、音響モデル管理サーバ２０，言語モデル管理サーバ３０において、音響データ取得手段１０６，言語データ取得手段１１３により、更新された音響データ１０７，更新された言語データ１１４を取得し、音響モデル構築手段１０８，言語モデル構築手段１１５により最新の音響モデル、最新の言語モデルを構築し、構築した最新の音響モデル、最新の言語モデルを、ネットワークを介して音声認識装置１０に送信し、音声認識装置１０において、音響モデル更新手段１１１，言語モデル更新手段１１８が、最新の音響モデル、最新の言語モデルにより、照合手段１０１が参照する音響モデル１０２，言語モデル１０３を更新することである。

【００７３】次に動作について説明する。音響モデル管理サーバ２０において、音響データ取得手段１０６は、音響モデル更新指令１０５と同期、あるいは非同期に動作し、常時、更新あるいは配信される音響データを自動、あるいは半自動的にダウンロードし、更新された音響データ１０７に格納する。取得するこれらの音響データは、例えばインターネット上で更新されたり、マルチメディア放送によって配信される音声データ、あるいは音声データと対応する書き起こしテキストであり、検索

ツールによってインターネット上で検索されたり、マルチメディア放送の番組表を用いて決定されてダウンロードされる。

【００７４】更新された音響データ１０７は、音響データ取得手段１０６により取得された音響モデル学習用音響データの集積であり、上記の例では、音声データや音声データと対応する書き起こしテキストからなる。

【００７５】音響モデル構築手段１０８は、例えば、一定の時間間隔、音声認識処理が実施された時間間隔、あるいは入力装置から与えられるユーザの指示等、適当なタイミングで与えられる音響モデル更新指令１０５を受けて、更新された音響データ１０７を参照し、統計的手法を用いて音響モデルのパラメータ推定を行うことによって、例えば、音声データのみからベクトル量子化アルゴリズムを用いることによって、あるいは音声データと対応する書き起こしテキストからバウム・ウェルチアルゴリズムを用いることによって、学習データを良く表すように音響モデルを構築して音響モデル格納手段１０９に格納する。

【００７６】音響モデル格納手段１０９は、音響モデル構築手段１０８により構築された音響モデルを記憶し、読み出し要求に応じて音響モデルを出力する。音響モデル送信手段１１０は、音響モデル格納手段１０９から音響モデルを読み出し、音声認識装置１０の音響モデル更新手段１１１に、ネットワークを介して送信する。

【００７７】音声認識装置１０において、音響モデル更新手段１１１は、音響モデル管理サーバ２０の音響モデル送信手段１１０からネットワークを介して受け取った音響モデルにより、照合手段１０１が照合の際に参照する音響モデル１０２を更新する。

【００７８】言語モデル管理サーバ３０において、言語データ取得手段１１３は、言語モデル更新指令１１２と同期、あるいは非同期に動作し、常時、更新あるいは配信される言語データをダウンロードし、言語モデル構築のために用いる新規の言語データを収集し、更新された言語データ１１４へ格納する。取得するこれらの言語データは、例えば、定期的に配信される新聞やメールマガジンやインターネット上から検索可能なテキスト、チャット、メール、マニュアル等のテキスト等である。

【００７９】更新された言語データ１１４は、言語データ取得手段１１３により取得された言語モデル学習用言語データの集積であり、テキストデータや同時に得られるテキスト内容に関するキーワード情報等である。

【００８０】言語モデル構築手段１１５は、例えば、一定の時間間隔、音声認識処理が実施された時間間隔、あるいは入力装置から与えられるユーザの指示等、適当なタイミングで与えられる言語モデル更新指令１１２を受けて、更新された言語データ１１４を参照してテキストデータを読み出し、統計的手法を用いて言語モデルのパラメータ推定を行うことによって、例えば単語に分割したテキストデータからｎグラム統計量を求めることによって、学習用データを良く表すように言語モデルを構築し、言語モデル格納手段１１６に格納する。

【００８１】言語モデル格納手段１１６は、言語モデル構築手段１１５により構築された言語モデルを記憶し、読み出し要求に応じて言語モデルを出力する。言語モデル送信手段１１７は、言語モデル格納手段１１６から言語モデルを読み出し、ネットワークを介して音声認識装置１０の言語モデル更新手段１１８へ送信する。

【００８２】音声認識装置１０において、言語モデル更新手段１１８は、言語モデル管理サーバ３０の言語モデル送信手段１１７からネットワークを介して受け取った言語モデルにより、照合手段１０１が照合の際に参照する言語モデル１０３を更新する。

【００８３】図２はこの発明の実施の形態１による音響モデル１０２の更新処理を示すフローチャートである。ステップＳＴ２０１において、音響モデル更新タイミング決定手段（図示していない）は、例えばユーザの指示や音響モデルの最終更新時刻からの時間間隔、ネットワークの利用状況等の監視から適当な更新タイミングを判定し、音響モデル管理サーバ２０の音響モデル構築手段１０８へ音響モデル更新指令１０５を送信する。音響モデル構築手段１０８は、音響モデル更新指令１０５を受けていれば、ステップＳＴ２０２へ進み、音響モデル更新指令１０５を受けていなければ処理を終了する。

【００８４】ステップＳＴ２０２において、音響モデル構築手段１０８は、学習に用いる更新された音響データ１０７を読み出す。ステップＳＴ２０３において、音響モデル構築手段１０８は、更新された音響データ１０７から、統計的手法を用いて音響モデルのパラメータ推定を行うことによって音響モデルを構築し、構築した音響モデルを音響モデル格納手段１０９へ格納する。

【００８５】ステップＳＴ２０４において、音響モデル送信手段１１０は、音響モデル格納手段１０９から音響モデルを読み出して、ネットワークを介して音声認識装置１０の音響モデル更新手段１１１へ音響モデルを送信する。ステップＳＴ２０５において、音響モデル更新手段１１１は、受け取った音響モデルにより照合手段１０１が参照する音響モデル１０２を更新する。

【００８６】なお、音響モデル更新指令１０５により、音響モデル１０２の更新を要求する際に、同時に音声認識装置１０が利用している音響モデル１０２のバージョンを伝達し、音響モデル送信手段１１０が、音響モデル格納手段１０９に格納されている音響モデル全体ではなく、それまで音声認識装置１０が利用していた音響モデル１０２との差分情報のみを送信すれば、送信データを減らすことができ、ネットワークの負荷を軽減することができる。

【００８７】また、音響モデル構築手段１０８が、あらかじめ更新された音響モデルを構築しておき、音響モデ

ル更新指令１０５の要求にしたがって、音響モデルを音声認識装置１０に送信するような形態を取った場合でも同様に動作可能である。

【００８８】さらに、音声認識装置１０がユーザ辞書を持つ場合であっても、同様に処理可能である。

【００８９】さらに、この実施の形態では、音声認識を対象として説明を行ったが、パターンとシンボルの関係を表した確率モデル、シンボルの出現を表した確率モデルからなるパターン認識を対象とするものであれば、同様に適用可能である。

【００９０】さらに、更新された音響データ１０７の格納形式は、音響モデル構築時に利用可能な形式であれば、あらかじめ信号処理や頻度分布を計算してあってもかまわない。

【００９１】図３はこの発明の実施の形態１による言語モデル１０３の更新処理を示すフローチャートである。ステップＳＴ３０１において、言語モデル更新タイミング決定手段（図示していない）は、例えば、ユーザの指示や言語モデルの最終更新時刻からの時間間隔、ネットワークの利用状況等の監視から適当な更新タイミングを判定し、言語モデル管理サーバ３０の言語モデル構築手段１１５に言語モデル更新指令１１２を送信する。言語モデル構築手段１１５は、言語モデル更新指令１１２を受けていればステップＳＴ３０２へ進み、言語モデル更新指令１１２を受けていなければ処理を終了する。

【００９２】ステップＳＴ３０２において、言語モデル構築手段１１５は、学習に用いる更新された言語データ１１４を読み出す。ステップＳＴ３０３において、言語モデル構築手段１１５は、更新された言語データ１１４から、統計的手法を用いて言語モデルのパラメータ推定を行うことによって言語モデルを構築し、構築した言語モデルを言語モデル格納手段１１６へ格納する。

【００９３】ステップＳＴ３０４において、言語モデル送信手段１１７は、言語モデル格納手段１１６から言語モデルを読み出して、ネットワークを介して音声認識装置１０の言語モデル更新手段１１８に送信する。ステップＳＴ３０５において、言語モデル更新手段１１８は、受け取った言語モデルにより照合手段１０１が参照する言語モデル１０３を更新する。

【００９４】なお、言語モデル更新指令１１２により、言語モデルの更新を要求する際に、同時に音声認識装置１０が利用している言語モデル１０３のバージョンを伝達することにより、言語モデル格納手段１１６に格納されている言語モデル全体ではなく、それまで音声認識装置１０が利用していた言語モデル１０３との差分情報のみを送信すれば、送信データを減らすことができ、ネットワークの負荷を軽減することができる。

【００９５】また、言語モデル構築手段１１５が、あらかじめ更新された言語モデルを構築しておき、言語モデル更新指令１１２の要求にしたがって、言語モデルを送信するような形態を取った場合でも、同様に動作可能である。

【００９６】さらに、音声認識装置１０がユーザ辞書を持つ場合であっても、同様に動作可能である。

【００９７】さらに、この実施の形態では、音声認識を対象として説明を行ったが、パターンとシンボルの関係を表した確率モデル、シンボルの出現を表した確率モデルからなるパターン認識を対象とするものであれば、同様に適用可能である。

【００９８】さらに、更新された言語データ１１４の格納形式は、言語モデル構築時に利用可能な形式であれば、あらかじめ単語に分割しておいたり、言語モデル構築に使えるように単語や単語連鎖、同時に出現する単語の組み合わせ等について頻度あるいは確率計算してあってもかまわない。

【００９９】この実施の形態１の図１では、音響モデル更新指令１０５に従って音響モデルを構築する場合について示しているが、音響データ取得手段１０６が音響データを取得した際に、音響モデル構築手段１０８が音響モデルを更新して音響モデル格納手段に１０９に格納し、音響モデル送信手段１１０が音響モデル更新指令１０５を受けて音響モデルを読み出すようにしても良い。言語モデルの更新についても同様である。

【０１００】また、図１では音響モデル更新手段１１１及び言語モデル更新手段１１８を備える場合を示したが、音響モデル更新手段１１１又は言語モデル更新手段１１８のどちらか一方のみを備える場合であってもかまわない。

【０１０１】なお、実施の形態１における音声認識システムを音声認識プログラムとして記録媒体に記録することもできる。この場合には、音響モデル管理サーバ２０において、音響データ取得手段１０６と同様の処理を行う音響データ取得機能と、音響モデル構築手段１０８と同様の処理を行う音響モデル構築機能と、音響モデル格納手段１０９と同様の処理を行う音響モデル格納機能と、音響モデル送信手段１１０と同様の処理を行う音響モデル送信機能から構成されるソフトウェアと、言語モデル管理サーバ３０において、言語データ取得手段１１３と同様の処理を行う言語データ取得機能と、言語モデル構築手段１１５と同様の処理を行う言語モデル構築機能と、言語モデル格納手段１１６と同様の処理を行う言語モデル格納機能と、言語モデル送信手段１１７と同様の処理を行う言語モデル送信機能から構成されるソフトウェアと、音声認識装置１０において、音響モデル更新手段１１１と同様の処理を行う音響モデル更新機能と、言語モデル更新手段１１８と同様の処理を行う言語モデル更新機能と、照合手段１０１と同様の処理を行う照合機能から構成されるソフトウェアで音声認識プログラムとなる。

【０１０２】記録媒体に記録する音声認識プログラム

は、音声認識装置１０のソフトウェアと、音響モデル管理サーバ２０のソフトウェアを、別々の記録媒体に記録しても良いし、１つの記録媒体に記録して、音声認識装置１０又は音響モデル管理サーバ２０から、それぞれ音響モデル管理サーバ２０又は音声認識装置１０に送信しても良い。また、これは言語モデルを対象とした場合でも同様である。

【０１０３】以上のように、この実施の形態１によれば、ネットワークに接続された音響モデル管理サーバ２０又は言語モデル管理サーバ３０で、更新された音響データ１０７又は更新された言語データ１１４を取得し、最新の状態にある音響モデル又は言語モデルを構築し、ネットワークを介してユーザ側の音声認識装置１０の音響モデル１０２又は言語モデル１０３を更新することで、ユーザに大きな負担をかけることなく、音声認識の認識精度を向上させることができるという効果が得られる。

【０１０４】実施の形態２．図４はこの発明の実施の形態２による音声認識システムの構成を示すブロック図である。図４の言語モデル管理サーバ３０において、４０１は更新する音声認識装置１０の言語モデル１０３を特定するＩＤを取得する更新言語モデルＩＤ取得手段、４０２は、更新言語モデルＩＤ取得手段４０１が取得したＩＤで指示される特定条件に対応して、更新された言語データ１１４を読み出す特定向け言語データ読み出し手段、４０３は、特定向け言語データ読み出し手段４０２により読み出された更新された言語データ１１４を参照し、特定条件に対応した言語モデルを構築する特定向け言語モデル構築手段である。既に説明した各手段及び各モデルについては、同一の符号を付し説明を省略する。

【０１０５】従来技術と異なるこの実施の形態に特徴的な部分は、更新言語モデルＩＤ取得手段４０１と、特定向け言語データ読み出し手段４０２と、特定向け言語モデル構築手段４０３とを備え、ネットワークを介して言語モデルＩＤにより特定された言語モデルを提供することである。ここで、特定向け言語モデルとは、特定のユーザやグループ、応用アプリケーション等に言語モデルを特化させることによって、より高い性能が得られるように学習した言語モデルである。

【０１０６】次に動作について説明する。言語管理サーバ３０において、更新言語モデルＩＤ取得手段４０１は、更新された複数の言語モデル１０３から特定の言語モデル１０３を選択するために用いられるＩＤを取得する。このＩＤは、例えば利用者のユーザＩＤ，音声信号１００の対象となるタスクを表すＩＤ等であり、更新する特定向けの言語モデル１０３を一意に定めることができるものである。

【０１０７】特定向け言語データ読み出し手段４０２は、更新言語モデルＩＤ取得手段４０１が取得した更新言語モデルＩＤを受け取り、更新された言語データ１１４を文や独立したテキストの単位で読み出して、例えば言語データに付与されるテキスト内容に関するキーワードや言語データに含まれるキーワードから判定し、言語モデルＩＤで特定される対象であるかどうか識別するフラグを付与し、特定向け言語モデル構築手段４０３へ送る。

【０１０８】特定向け言語モデル構築手段４０３は、更新された言語データ１１４から特定対象について認識精度が高くなるように学習した言語モデルを構築し、言語モデル格納手段１１６に格納する。

【０１０９】このために、まず、学習用言語データを文、あるいは複数の文を単位として含まれるキーワード等から、特定される言語モデル１０３を判定しておき、これに従って、例えば「対話音声認識のための事前タスク適応の検討」、伊藤彰則、好田正紀、電子情報通信学会技術研究報告（ＳＰ９６－８１）、１９９６年（以下、文献４とする）で検討されているように、関連深いテキストデータにより大きな重みを付与することによって、特定条件に対応した言語モデルを構築できる。

【０１１０】例えば、スポーツに関するトピックに特化した特定向け言語モデルを構築するには、言語モデル学習時に、特定向け言語データ読み出し手段４０２から得たフラグを参照し、スポーツに関するトピックのテキストデータであれば、実際の頻度をα倍して数え、それ以外の記事であれば、そのまま頻度で両者を加えて確率モデルを推定する。ここで、αは音声認識の対象とする特定向けテキストのうち、学習に用いないデータに対する言語モデルのエントロピーが最小となるように定める。

【０１１１】図５はこの発明の実施の形態２による言語モデル１０３の更新処理を示すフローチャートである。ステップＳＴ５０１において、言語モデル更新タイミング決定手段（図示していない）は、例えば、ユーザの指示や最終更新時刻からの時間間隔、ネットワークの利用状況等の監視から適当な更新タイミングを判定し、言語モデル管理サーバ３０の更新言語モデルＩＤ取得手段４０１へ言語モデル更新指令１１２を送信する。更新言語モデルＩＤ取得手段４０１は、言語モデル更新指令１１２を受けていれば、ステップＳＴ５０２へ進み、言語モデル更新指令１１２を受けていなければ、処理を終了する。

【０１１２】ステップＳＴ５０２において、更新タイミングであれば、更新言語モデルＩＤ取得手段４０１は、使用しているユーザ・グループを特定する手段、タスクを特定する手段等により、更新先の言語モデル１０３のＩＤを取得し、特定向け言語データ読み出し手段４０２へ送る。ステップＳＴ５０３において、特定向け言語データ読み出し手段４０２は、特定向け言語モデルＩＤに従い、更新された言語データ１１４を文や独立したテキストの単位で読み出して、言語モデルＩＤで特定される対象であるかどうか判別するフラグを付与して、更新さ

れた言語データ１１４を読み出す。

【０１１３】ステップＳＴ５０４において、特定向け言語モデル構築手段４０３は、学習アルゴリズムに従い、更新された言語データ１１４から言語モデルを推定し、推定した言語モデルを言語モデル格納手段１１６へ格納する。ステップＳＴ５０５において、言語モデル送信手段１１７は、読み出した言語モデルを、ネットワークを介して音声認識装置１０の言語モデル更新手段１１８に送信する。ステップＳＴ５０６において、言語モデル更新手段１１８は、受け取った言語モデルにより照合手段１０１が参照する言語モデル１０３を更新する。

【０１１４】なお、言語モデルＩＤに対して、適合する言語モデルを出力することができれば、特定向け言語モデルをあらかじめ構築しておく必要はない。例えば、言語モデルＩＤに依存して学習用言語データのみ作成しておき、要求に応じて言語モデルを構築しても良い。

【０１１５】なお、この実施の形態では、文献４にしたがった特定向け言語モデル構成法を例としたが、言語モデルを決定できるＩＤを用いて、複数の言語モデル１０３から選択する方法であれば同様に適用可能である。

【０１１６】また、言語モデル更新指令１１２により、言語モデル更新要求の際に、同時に音声認識装置１０が利用している言語モデル１０３のバージョンを伝達することにより、構築した特定向け言語モデル全体ではなく、現行の言語モデル１０３からの差分情報のみを送信し、ネットワークの負荷を軽減することができる。

【０１１７】さらに、音声認識装置にユーザ辞書６０１がある場合であっても、同様に処理可能である。

【０１１８】さらに、この説明では音声認識を対象として説明を行ったが、パターンとシンボルの関係を表した確率モデル、シンボルの出現を表した確率モデルからなるパターン認識を対象とするものであれば同様に適用可能である。

【０１１９】また、この実施の形態では、特定向け言語モデルを構築し、構築した特定向け言語モデルにより言語モデル１０３を更新しているが、特定向け音響モデルを構築し、構築した特定向け音響モデルにより音響モデル１０２を更新することもできる。その場合には、図４において、言語モデル管理サーバ３０の代わりに音響モデル管理サーバ，言語データ取得手段１１３の代わりに音響データ取得手段、更新された言語データ１１４の代わりに更新された音響データ、更新言語モデルＩＤ取得手段４０１の代わりに更新音響モデルＩＤ取得手段、特定向け言語データ読み出し手段４０２の代わりに特定向け音響モデル読み出し手段、特定向け言語モデル構築手段４０３の代わりに特定向け音響モデル構築手段、言語モデル格納手段１１６の代わりに音響モデル格納手段、言語モデル送信手段１１７の代わりに音響モデル送信手段を備え、音声認識装置１０において、言語モデル更新手段１１８の代わりに音響モデル更新手段を備え、音響モデル更新手段が音響モデル１０２を更新するようにすれば良い。

【０１２０】さらに、実施の形態２における音声認識システムを音声認識プログラムとして記録媒体に記録することもできる。この場合には、言語モデル管理サーバ３０において、言語データ取得手段１１３と同様の処理を行う言語データ取得機能と、更新言語モデルＩＤ取得手段４０１と同様の処理を行う更新言語モデルＩＤ取得機能と、特定向け言語データ読み出し手段４０２と同様の処理を行う特定向け言語データ読み出し機能と、特定向け言語モデル構築手段４０３と同様の処理を行う特定向け言語モデル構築機能と、言語モデル格納手段１１６と同様の処理を行う言語モデル格納機能と、言語モデル送信手段１１７と同様の処理を行う言語モデル送信機能から構成されるソフトウェアと、音声認識装置１０において、言語モデル更新手段１１８と同様の処理を行う言語モデル更新機能と、照合手段１０１と同様の処理を行う照合機能から構成されるソフトウェアで音声認識プログラムとなる。これは音響モデルを対象とした場合でも同様である。

【０１２１】記録媒体に記録する音声認識プログラムは、音声認識装置１０のソフトウェアと、言語モデル管理サーバ３０のソフトウェアを、別々の記録媒体に記録しても良いし、１つの記録媒体に記録して、音声認識装置１０又は言語モデル管理サーバ３０から、それぞれ言語モデル管理サーバ３０又は音声認識装置１０に送信しても良い。また、これは音響モデルを対象とした場合でも同様である。

【０１２２】以上のように、この実施の形態２によれば、ネットワークに接続された言語モデル管理サーバ３０で、更新された言語データ１１４を取得し、しかも、ユーザの音声認識装置１０の言語モデル１０３のＩＤを取得することで、最新の状態にあり、しかもユーザに対応した特定向けの言語モデルを構築し、ネットワークを介してユーザの音声認識装置１０の言語モデル１０３を更新することで、ユーザに大きな負担をかけることなく、音声認識の認識精度を向上させることができるという効果が得られる。

【０１２３】また、この実施の形態２によれば、特定向けにカスタマイズされた言語モデルを、ネットワークを介して更新することにより、ユーザが複数の異なる照合手段１０１を利用する場合でも、全ての照合手段１０１の利用時に適切な言語モデル１０３を利用し、高い認識精度を得ることができるという効果が得られる。

【０１２４】実施の形態３．図６はこの発明の実施の形態３による音声認識システムの構成を示すブロック図である。図６の音声認識装置１０において、６０１は照合手段１０１が照合の際に参照する単語を登録したユーザ辞書であり、言語モデル管理サーバ３０において、６０２は言語モデル更新指令１１２を受けて、照合手段１０

１が参照するユーザ辞書６０１を、ネットワークを介して読み出すユーザ辞書読み出し手段、６０３は、更新された言語データ１１４とユーザ辞書読み出し手段６０２が読み出したユーザ辞書６０１を参照し、ユーザ辞書６０１に依存した言語モデルを構築するユーザ辞書依存言語モデル構築手段である。既に説明した各手段及び各モデルについては、同一の符号を付し説明を省略する。

【０１２５】従来技術と異なるこの実施の形態に特徴的な部分は、ユーザ辞書読み出し手段６０２とユーザ辞書依存言語モデル構築手段６０３を備えたことである。

【０１２６】次に動作について説明する。言語モデル管理サーバ３０のユーザ辞書読み出し手段６０２は、言語モデル更新指令１１２を受けて、音声認識装置１０の照合手段１０１が参照するユーザ辞書６０１を、ネットワークを介して読み出す。ユーザ辞書依存言語モデル構築手段６０３は、ユーザ辞書６０１に登録された単語と更新された言語データ１１４を用いて、最新の状態に更新されており、かつユーザにカスタマイズされた言語モデルを構築する。

【０１２７】ユーザ辞書６０１に依存した言語モデルの構築は、例えば、更新された言語データ１１４の中から、ユーザ辞書６０１に存在する単語が含まれるテキストを抜き出し、これを特定向けテキストであると見なして、実施の形態２で参照した文献４記載の方法を実施することにより行われる。これによって、ユーザ辞書６０１記載の単語のうち、元の言語モデルでは登録されておらず、適切な統計量が付与されていなかった単語で、更新されたテキストにおいて出現した単語に妥当な統計量を付与することが可能となり、認識精度が向上することを期待できる。

【０１２８】図７はこの発明の実施の形態３による言語モデル１０３の更新処理を示すフローチャートである。ステップＳＴ７０１において、言語モデル更新タイミング決定手段（図示していない）は、例えば、ユーザの指示や最終更新時刻からの時間間隔、ネットワークの利用状況等の監視から適当な更新タイミングを判定し、言語モデル管理サーバ３０のユーザ辞書読み出し手段６０２へ言語モデル更新指令１１２を送信する。ユーザ辞書読み出し手段６０２は、言語モデル更新指令１１２を受けていれば、ステップＳＴ７０２へ進み、言語モデル更新指令１１２を受けていなければ、処理を終了する。

【０１２９】ステップＳＴ７０２において、ユーザ辞書読み出し手段６０２は、照合手段１０１が参照するユーザ辞書６０１を、ネットワークを介して読み出す。ステップＳＴ７０３において、ユーザ辞書依存言語モデル構築手段６０３は、さらに更新された言語データ１１４を読み出す。ステップＳＴ７０４において、ユーザ辞書依存言語モデル構築手段６０３は、ユーザ辞書６０１及び更新された言語データ１１４からユーザ辞書６０１に依存した言語モデルを構築し、言語モデル格納手段１１６に格納する。

【０１３０】ステップＳＴ７０５において、言語モデル送信手段１１７は、言語モデル格納手段１１６から読み出したユーザ辞書６０１に依存した言語モデルを、ネットワークを介して音声認識装置１０の言語モデル更新手段１１８に送信する。ステップＳＴ７０６において、言語モデル更新手段１１８は、受け取った言語モデルにより照合手段１０１が参照する言語モデル１０３を更新する。

【０１３１】また、この実施の形態では、音声認識を対象として説明を行ったが、パターンとシンボルの関係を表した確率モデル、シンボルの出現を表した確率モデルからなるパターン認識を対象とするものであれば、同様に適用可能である。

【０１３２】さらに、実施の形態３における音声認識システムを音声認識プログラムとして記録媒体に記録することもできる。この場合には、言語モデル管理サーバ３０において、言語データ取得手段１１３と同様の処理を行う言語データ取得機能と、ユーザ辞書読み出し手段６０２と同様の処理を行うユーザ辞書読み出し機能と、ユーザ辞書依存言語モデル構築手段６０３と同様の処理を行うユーザ辞書依存言語モデル構築機能と、言語モデル格納手段１１６と同様の処理を行う言語モデル格納機能と、言語モデル送信手段１１７と同様の処理を行う言語モデル送信機能から構成されるソフトウェアと、音声認識装置１０において、言語モデル更新手段１１８と同様の処理を行う言語モデル更新機能と、照合手段１０１と同様の処理を行う照合機能から構成されるソフトウェアで音声認識プログラムとなる。

【０１３３】記録媒体に記録する音声認識プログラムは、音声認識装置１０のソフトウェアと、言語モデル管理サーバ３０のソフトウェアを、別々の記録媒体に記録しても良いし、１つの記録媒体に記録して、音声認識装置１０又は言語モデル管理サーバ３０から、それぞれ言語モデル管理サーバ３０又は音声認識装置１０に送信しても良い。

【０１３４】以上のように、この実施の形態３によれば、ネットワークに接続された言語モデル管理サーバ３０で、更新された言語データ１１４を取得し、しかも、ユーザの音声認識装置１０のユーザ辞書６０１を読み出すことで、最新の状態にあり、しかもユーザ辞書６０１に登録した単語について、より詳細に反映させた言語モデルを構築し、ネットワークを介してユーザの音声認識装置１０の言語モデル１０３を更新することで、ユーザ辞書が大きくなった場合でも、ユーザに大きな負担をかけることなく、音声認識の認識精度を向上させることができるという効果が得られる。

【０１３５】実施の形態４．図８はこの発明の実施の形態４による音声認識システムの構成を示すブロック図である。図８の言語モデル管理サーバ３０において、８０

１は、言語モデル更新指令１１２を受けて、ユーザが利用したテキストを取得するユーザ利用テキスト取得手段、８０２はユーザ利用テキスト取得手段８０１が取得したテキストを格納するユーザ利用テキスト格納手段、８０３は、更新された言語データ１１４と、ユーザ利用テキスト格納手段８０２に格納されているテキストを参照し、テキストに依存した言語モデルを構築するユーザ利用テキスト依存言語モデル構築手段である。既に説明した各手段及び各モデルについては、同一の符号を付し説明を省略する。

【０１３６】従来技術と異なるこの実施の形態に特徴的な部分は、ユーザ利用テキスト取得手段８０１，ユーザ利用テキスト格納手段８０２及びユーザ利用テキスト依存言語モデル構築手段８０３を備え、ユーザが利用したテキストと最新状態に更新された言語データ１１４を参照し、ユーザが利用したテキストに合わせて言語モデルを構築することである。

【０１３７】次に動作について説明する。言語モデル管理サーバ３０のユーザ利用テキスト取得手段８０１は、言語モデル更新指令１１２を受けて、例えば、ユーザがあらかじめ指定したファイル、ディレクトリを走査することにより、ユーザが参照、あるいは記述したテキストファイルを読み出す。ユーザ利用テキスト格納手段８０２は、ユーザ利用テキスト取得手段８０１によって収集されたテキストを格納する。

【０１３８】ユーザ利用テキスト依存言語モデル構築手段８０３は、ユーザ利用テキスト及び更新された言語データ１１４を参照し、認識精度が高くなるように言語モデルを構築する。ユーザ利用テキストを用いた言語モデルの構築では、例えば、ユーザ利用テキストを特定向けテキストであると見なし、実施の形態２で参照した文献４記載の方法を実施することにより、ユーザ利用テキスト依存の言語モデルを構築する。このようにして構築された言語モデルは、ユーザが参照あるいは既出したテキストの性質を反映させているため、ユーザが発声する確率の高い言語的性質を含み、より精度の高い認識結果を得ることができる。

【０１３９】図９はこの発明の実施の形態４による言語モデル１０３の更新処理を示すフローチャートである。ステップＳＴ９０１において、言語モデル更新タイミング決定手段（図示していない）は、ユーザの指示や最終更新時刻からの時間間隔、ネットワークの利用状況等のモニタから適当なタイミングを判定し、言語モデル管理サーバ３０のユーザ利用テキスト取得手段８０１へ言語モデル更新指令１１２を送信する。ユーザ利用テキスト取得手段８０１は、言語モデル更新指令１１２を受けていれば、ステップＳＴ９０２へ進み、言語モデル更新指令１１２を受けていなければ、処理を終了する。

【０１４０】ステップＳＴ９０２において、ユーザ利用テキスト取得手段８０１は、ユーザ利用テキストを読み出しユーザ利用テキスト格納手段８０２へ格納する。ステップＳＴ９０３において、ユーザ利用テキスト依存言語モデル構築手段８０３は、ユーザ利用テキストと更新された言語データ１１４を読み出す。ステップＳＴ９０４において、ユーザ利用テキスト依存言語モデル構築手段８０３は、ユーザ利用テキスト及び更新された言語データ１１４からユーザ利用テキスト依存言語モデルを構築し言語モデル格納手段１１６に格納する。

【０１４１】ステップＳＴ９０５において、言語モデル送信手段１１７は、言語モデル格納手段１１６から読み出したユーザ利用テキスト依存言語モデルを、ネットワークを介して音声認識装置１０の言語モデル更新手段１１８に送信する。ステップＳＴ９０６において、言語モデル更新手段１１８は、受け取った言語モデルにより照合手段１０１が参照する言語モデル１０３を更新する。

【０１４２】この実施の形態では、ユーザ利用テキストの入手を、特定ディレクトリやファイルの検索によるとしたが、ユーザ利用テキスト取得手段８０１は、テキストが収集できるのであれば、テキストをファイルやディレクトリから取り出すのではなく、音声認識やキーボード、ペン、ＯＣＲ等のユーザ入力、あるいはブラウザ等によってユーザが閲覧したテキスト等を利用してもかまわない。

【０１４３】また、ユーザ利用テキスト格納手段８０２は、ユーザ利用テキスト取得手段８０１によって収集されたテキストを格納するとしたが、ユーザ利用テキスト依存言語モデル構築手段８０３の基準に従い、テキストを適当な手段によって単語に分割したり、モデル構築時に参照する単語、単語連鎖、同時に出現する単語の組み合わせ等に関する頻度として格納しても同様である。

【０１４４】さらに、この実施の形態では、音声認識を対象として説明を行ったが、パターンとシンボルの関係を表した確率モデル、シンボルの出現を表した確率モデルからなるパターン認識を対象とするものであれば同様に適用可能である。

【０１４５】さらに、実施の形態４における音声認識システムを、音声認識プログラムとして記録媒体に記録することもできる。この場合には、言語モデル管理サーバ３０において、言語データ取得手段と同様の処理を行う言語データ取得機能と、ユーザ利用テキスト取得手段８０１と同様の処理を行うユーザ利用テキスト取得機能と、ユーザ利用テキスト格納手段８０２と同様の処理を行うユーザ利用テキスト格納機能と、ユーザ利用テキスト依存言語モデル構築手段８０３と同様の処理を行うユーザ利用テキスト依存言語モデル構築機能と、言語モデル格納手段１１６と同様の処理を行う言語モデル格納と、言語モデル送信手段１１７と同様の処理を行う言語モデル送信機能から構成されるソフトウェアと、音声認識装置１０において、言語モデル更新手段１１８と同様の処理を行う言語モデル更新機能と、照合手段１０１と

同様の処理を行う照合機能から構成されるソフトウェアで音声認識プログラムとなる。

【０１４６】記録媒体に記録する音声認識プログラムは、音声認識装置１０のソフトウェアと、言語モデル管理サーバ３０のソフトウェアを、別々の記録媒体に記録しても良いし、１つの記録媒体に記録して、音声認識装置１０又は言語モデル管理サーバ３０から、それぞれ言語モデル管理サーバ３０又は音声認識装置１０に送信しても良い。

【０１４７】以上のように、この実施の形態４によれば、ネットワークに接続された言語モデル管理サーバ３０で、更新された言語データ１１４を取得し、しかも、ユーザが利用したテキストを取得することで、最新の状態にあり、しかもユーザが利用したテキストに依存した言語モデルを構築し、ネットワークを介してユーザの音声認識装置１０の言語モデル１０３を更新することで、ユーザに大きな負担をかけることなく、音声認識の認識精度を向上させることができるという効果が得られる。

【０１４８】実施の形態５．図１０はこの発明の実施の形態５による音声認識システムの構成を示すブロック図である。図１０の音声認識装置１０において、１００１は照合手段１０１が照合の際に参照する音響モデルを識別するＩＤを取得する音響モデルＩＤ取得手段、１００２は、音響モデルＩＤ取得手段１００１が取得したＩＤを読み込み、入力された音声信号から適応化用の音声データを取得し、読み込んだＩＤと取得した適応化用の音声データを、ネットワークを介して音響モデル管理サーバ２０に送信する適応化用音声取得手段である。

【０１４９】図１０の音響モデル管理サーバ２０において、１００３は適応化前の初期音響モデル、１００４は、音声認識装置１０の適応化用音声取得手段１００２から送信された適応化用の音声データを用いて、適応化前の初期音響モデル１００３を適応化し、適応化済み音響モデルを、適応化用音声取得手段１００２から送信されたＩＤに対応付けて適応化済み音響モデル格納手段１００５に格納する音響モデル適応化手段、１００６は、音響モデル更新指令１０５を受けて、ネットワークを介して音声認識装置１０の音響モデルＩＤ取得手段１００１が取得したＩＤを受信し、受信したＩＤに対応する適応化済み音響モデルを、適応化済み音響モデル格納手段１００５から選択して読み出す適応化済み音響モデル選択手段である。既に説明した各手段及び各モデルについては、同一の符号を付し説明を省略する。

【０１５０】従来技術と異なるこの実施の形態に特徴的な部分は、音声認識装置１０において、音響モデルＩＤ取得手段１００１，適応化用音声取得手段１００２を備え、音響モデル管理サーバ２０において、初期音響モデル１００３，音響モデル適応化手段１００４，適応化済み音響モデル格納手段１００５及び適応化済み音響モデル選択手段１００６を備えたことである。

【０１５１】この実施の形態では、音声認識装置１０において、適応化対象となる音響モデルＩＤと適応化用の音声データを取得し、音響モデル管理サーバ２０において、音響モデルＩＤに依存した適応化を行った音響モデルを構築して、音声認識装置１０に送信し、音声認識装置１０において、音響モデル１０２を更新することにより、ユーザは任意の照合手段１０１の利用に際して、適応化した音響モデルを参照可能であるため、より高い認識精度を得ることができる。

【０１５２】次に動作について説明する。音声認識装置１０において、音響モデルＩＤ取得手段１００１は、適応化対象となる音響モデルを決定するものであり、例えば、音声認識装置１０を利用するユーザのユーザＩＤである。適応化用音声取得手段１００２は、照合手段１０１の利用前にあらかじめ適応化用として音声信号１００による適応化用の音声データを取得し、音響モデルＩＤ取得手段１００１から読み出す音響モデルＩＤと、取得したした適応化用の音声データを、ネットワークを介して接続される音響モデル管理サーバ２０の音響モデル適応化手段１００４へ送信する。

【０１５３】音響モデル管理サーバ２０において、初期音響モデル１００３は適応化を行う前の音響モデルである。音響モデル適応化手段１００４は、ネットワークを介して受け取った適応化用の音声データと初期音響モデル１００３を用いて、適応化した音響モデルを構築し、適応化済みの音響モデルとネットワークを介して受け取った音響モデルＩＤを、適応化済み音響モデル格納手段１００５へ格納する。音響モデルの適応化には、例えば最大事後確率推定法を用いる。

【０１５４】適応化済み音響モデル格納手段１００５は、音響モデルＩＤと音響モデル適応化手段１００４により適応化された音響モデルを格納し、適応化済み音響モデル選択手段の要求に従い、指定の音響モデルＩＤを持つ音響モデルを出力する。適応化済み音響モデル選択手段１００６は、音響モデル更新指令１０５を受けて、音響モデルＩＤ取得手段１００１から音響モデルＩＤを取得し、対応する適応化済み音響モデルを、適応化済み音響モデル格納手段１００５から選択して読み出す。

【０１５５】図１１はこの発明の実施の形態５による音響モデル１０２の更新処理を示すフローチャートである。処理はステップＳＴ１１０１からＳＴ１１０７までの音響モデルの適応化段階と、ステップＳＴ１１０８からＳＴ１１１２までの音響モデルの更新段階に分けられる。適応化段階では、入力された適応化用の音声データを用いて、音響モデルＩＤに依存した音響モデルを構築する。

【０１５６】ステップＳＴ１１０１において、音声認識装置１０の音響モデルＩＤ取得手段１００１は、適応化対象となる音響モデルを識別する、例えばユーザ名等の識別情報を取得する。ステップＳＴ１１０２において、

適応化用音声取得手段１００２は、音響モデルＩＤを読み出し、音声信号１００から入力される適応化用の音声データを取得する。ステップＳＴ１１０３において、適応化用音声取得手段１００２は、ネットワークを介して音響モデル管理サーバ２０に音響モデルの適応化要求を送信し、同時に音響モデルＩＤと適応化用の音声データを音響モデル適応化手段１００４へ送信する。

【０１５７】ステップＳＴ１１０４において、音響モデル適応化手段１００４は、ネットワークを介して音響モデルの適応化要求を受信し、音響モデルＩＤと適応化用の音声データを読み出す。ステップＳＴ１１０５において、音響モデル適応化手段１００４は初期音響モデル１００３を読み出す。ステップＳＴ１１０６において、音響モデル適応化手段１００４は、ネットワークを介して受け取った適応化用の音声データを用いて初期音響モデル１００３を適応化する。ステップＳＴ１１０７において、音響モデル適応化手段１００４は、適応化された音響モデルを音響モデルＩＤによって区別できるように、適応化済み音響モデル格納手段１００５に格納する。

【０１５８】ステップＳＴ１１０８において、音響モデル更新タイミング決定手段（図示していない）は、ユーザの指示や最終更新時刻からの時間間隔、ネットワークの利用状況等の監視から適当な更新タイミングを判定し、音響モデル更新指令１０５を適応化済み音響モデル選択手段１００６へ送信する。適応化済み音響モデル選択手段１００６は、音響モデル更新指令１０５を受けていれば、ステップＳＴ１１０９へ進み、音響モデル更新指令１０５を受けていなければ処理を終了する。ステップＳＴ１１０９において、適応化済み音響モデル選択手段１００６は、音声認識装置１０の音響モデルＩＤ取得手段１００１から、ネットワークを介して適応化対象の音響モデルＩＤを読み出す。

【０１５９】ステップＳＴ１１１０において、適応化済み音響モデル選択手段１００６は、音響モデルＩＤで指定された音響モデルを、適応化済み音響モデル格納手段１００５から選択して読み出す。ステップＳＴ１１１１において、音響モデル送信手段１１０は、読み出された適応化済み音響モデルを、ネットワークを介して音声認識装置１０の音響モデル更新手段１１１に送信する。ステップＳＴ１１１２において、音響モデル更新手段１１１は、受け取った適応化済み音響モデルにより照合手段１０１が参照する音響モデル１０２を更新する。

【０１６０】なお、音声認識装置１０において、音響モデルＩＤ取得手段１００１は、適応化対象となる音響モデルを決定するものであれば、回線伝達特性、背景雑音特性、残響音特性等を決定するものであってもかまわない。

【０１６１】また、音声認識装置１０において、適応化用音声取得手段１００２は、照合手段１０１の利用前にあらかじめ適応化用としてユーザの音声データを取得しているが、照合手段１０１が照合時に音声データを取得してして適応化用音声取得手段１００２に入力することで、次の照合の際に参照する音響モデルの適応化を行うことも可能である。

【０１６２】さらに、音声認識装置１０において、適応化用の音声のデータの取得と音響モデルＩＤの取得の手順は逆であってもかまわない。

【０１６３】さらに、音声認識装置１０において、適応化用音声取得手段１００２の音声データの格納形態は音声波形としたが、音声データを信号処理した音響特徴ベクトルの時系列、音響特徴ベクトルとベクトル量子化コードブックを参照して得られるコードブック符号列、それらを統計処理して得られる頻度分布、頻度分布から得られる確率分布等、音響モデルの学習に利用できる形態であれば、どのようなものでもかまわない。

【０１６４】さらに、音響モデル管理サーバ２０において、音響モデル適応化手段１００４が受信した適応化用の音声データを格納する手段を追加し、格納した多くの適応化用の音声データにより初期モデル１００３を更新することにより、学習の精度が高くなり、より高精度な認識を行う適応化済み音響モデルを構築することができる。

【０１６５】さらに、この実施の形態では音声認識を対象としたが、パターンとシンボルの関係を表した確率モデル、シンボルの出現を表した確率モデルからなるパターン認識を対象とするものであれば同様に適用可能である。

【０１６６】さらに、実施の形態５における音声認識システムを音声認識プログラムとして記録媒体に記録することもできる。この場合には、音声認識装置１０において、音響モデルＩＤ取得手段１００１と同様の処理を行う音響モデルＩＤ取得機能と、適応化用音声取得手段１００２と同様の処理を行う適応化用音声取得機能と、音響モデル更新手段１１１と同様の処理を行う音響モデル更新機能と、照合手段１０１と同様の処理を行う照合機能から構成されるソフトウェアと、音響モデル管理サーバ２０において、音響モデル適応化手段１００４と同様の処理を行う音響モデル適応化機能と、適応化済み音響モデル格納手段１００５と同様の処理を行う適応化済み音響モデル格納機能と、適応化済み音響モデル選択手段１００６と同様の処理を行う適応化済み音響モデル選択機能と、音響モデル送信手段１１０と同様の処理を行う音響モデル送信機能から構成されるソフトウェアで音声認識プログラムとなる。

【０１６７】記録媒体に記録する音声認識プログラムは、音声認識装置１０のソフトウェアと、音響モデル管理サーバ２０のソフトウェアを、別々の記録媒体に記録しても良いし、１つの記録媒体に記録して、音声認識装置１０又は音響モデル管理サーバ２０から、それぞれ音響モデル管理サーバ２０又は音声認識装置１０に送信し

ても良い。

【０１６８】以上のように、この実施の形態５によれば、ネットワークに接続された音響言語モデル管理サーバ２０で、ユーザの音声信号１００による適応化用の音声データにより適応化済みの音響モデルを構築し、ネットワークを介してユーザの音声認識装置１０の音響モデル１０２を更新することで、ユーザに大きな負担をかけることなく、音声認識の認識精度を向上させることができるという効果が得られる。

【０１６９】また、この実施の形態５によれば、ユーザに適応化した適応化済み音響モデルにより、ネットワークを介して音響モデル１０２を更新することで、ユーザが複数の異なる照合手段１０１を利用する場合でも、全ての照合手段１０１の利用時に適切な音響モデル１０２を利用し、高い認識精度を得ることができるという効果が得られる。

【０１７０】

【発明の効果】以上のように、この発明によれば、音響モデル管理サーバが、更新された音響データを取得して構築した音響モデルを、ネットワークを介して音声認識装置に送信し、音声認識装置が、音声認識の際に参照する音響モデルを、音響モデル管理サーバが送信した音響モデルにより更新することにより、ユーザに大きな負担をかけることなく、音声認識の認識精度を向上させることができるという効果がある。

【０１７１】この発明によれば、音響モデル管理サーバが、音声認識装置が音声認識の際に参照する音響モデルを特定するＩＤを取得し、取得したＩＤで指示される特定条件に対応して、更新された音響データを読み出し、特定条件に依存した音響モデルを構築して音声認識装置に送信することにより、ユーザに大きな負担をかけることなく、音声認識の認識精度を向上させることができると共に、特定向けにカスタマイズされた音響モデルを、ネットワークを介して送信することにより、ユーザが複数の異なる音声認識装置を利用する場合でも適切な音響モデルを利用し、高い認識精度を得ることができるという効果がある。

【０１７２】この発明によれば、言語モデル管理サーバが、更新された言語データを取得して構築した言語モデルを、ネットワークを介して音声認識装置に送信し、音声認識装置が、音声認識の際に参照する言語モデルを、言語モデル管理サーバが送信した言語モデルにより更新することにより、ユーザに大きな負担をかけることなく、音声認識の認識精度を向上させることができるという効果がある。

【０１７３】この発明によれば、言語モデル管理サーバが、音声認識装置が音声認識の際に参照する言語モデルを特定するＩＤを取得し、取得したＩＤで指示される特定条件に対応して、更新された言語データを読み出し、特定条件に依存した言語モデルを構築して音声認識装置に送信することにより、ユーザに大きな負担をかけることなく、音声認識の認識精度を向上させることができると共に、特定向けにカスタマイズされた言語モデルを、ネットワークを介して更新することにより、ユーザが複数の異なる音声認識装置を利用する場合でも適切な言語モデルを利用し、高い認識精度を得ることができるという効果がある。

【０１７４】この発明によれば、音声認識装置が音声認識の際に単語を登録したユーザ辞書を参照し、言語モデル管理サーバが、ネットワークを介してユーザ辞書を読み出し、更新された言語データと、読み出したユーザ辞書とを参照し、ユーザ辞書に依存した言語モデルを構築して音声認識装置に送信することにより、ユーザ辞書が大きくなった場合でも、ユーザに大きな負担をかけることなく、音声認識の認識精度を向上させることができるという効果がある。

【０１７５】この発明によれば、言語モデル管理サーバが、音声認識装置のユーザが利用したテキストを取得し、更新された言語データと、取得したテキストとを参照し、テキストに依存した言語モデルを構築して上記音声認識装置に送信することにより、ユーザに大きな負担をかけることなく、音声認識の認識精度を向上させることができるという効果がある。

【０１７６】この発明によれば、音声認識装置が、音響モデルを特定するＩＤと、入力された音声信号から適応化用の音声データとを取得し、取得したＩＤ及び適応化用の音声データを、ネットワークを介して音響モデル管理サーバに送信し、音響モデル管理サーバが、送信された適応化用の音声データを用いて、初期音響モデルを適応化し、適応化済み音響モデルを、送信されたＩＤに対応付けて格納すると共に、外部からの音響モデル更新指令を受けて、ネットワークを介して音声認識装置からＩＤを受信し、受信したＩＤに対応する適応化済み音響モデルを、格納している適応化済み音響モデルの中から選択して読み出し、ネットワークを介して音声認識装置に送信し、音声認識装置が、音声認識の際に参照する音響モデルを、音響モデル管理サーバが送信した適応化済み音響モデルにより更新することにより、ユーザに大きな負担をかけることなく、音声認識の認識精度を向上させることができると共に、ユーザに適応化した適応化済み音響モデルにより、ネットワークを介して音響モデルを更新することで、ユーザが複数の異なる音声認識装置を利用する場合でも、適切な音響モデルを利用し、高い認識精度を得ることができるという効果がある。

【０１７７】この発明によれば、音声認識装置がネットワークを介して接続された音響モデル管理サーバから、更新された音響データにより構築された音響モデルを受信し、照合手段が音声認識の際に参照する音響モデルを、受信した音響モデルにより更新する音響モデル更新手段とを備えたことにより、ユーザに大きな負担をかけ

ることなく、音声認識の認識精度を向上させることができるという効果がある。

【０１７８】この発明によれば、音響モデル更新手段が、ネットワークを介して接続された音響モデル管理サーバから、更新された音響データにより構築された、照合手段が音声認識の際に参照する音響モデルの特定条件に依存した音響モデルを受信し、照合手段が音声認識の際に参照する音響モデルを、受信した音響モデルにより更新することにより、ユーザに大きな負担をかけることなく、音声認識の認識精度を向上させることができると共に、特定向けにカスタマイズされた音響モデルを、ネットワークを介して受信することにより、ユーザが複数の異なる照合手段を利用する場合でも適切な音響モデルを利用し、高い認識精度を得ることができるという効果がある。

【０１７９】この発明によれば、音声認識装置がネットワークを介して接続された言語モデル管理サーバから、更新された言語データにより構築された言語モデルを受信し、照合手段が音声認識の際に参照する言語モデルを、受信した言語モデルにより更新する言語モデル更新手段とを備えたことにより、ユーザに大きな負担をかけることなく、音声認識の認識精度を向上させることができるという効果がある。

【０１８０】この発明によれば、言語モデル更新手段が、ネットワークを介して接続された言語モデル管理サーバから、更新された言語データにより構築された、照合手段が音声認識の際に参照する言語モデルの特定条件に依存した言語モデルを受信し、照合手段が音声認識の際に参照する言語モデルを、受信した言語モデルにより更新することにより、ユーザに大きな負担をかけることなく、音声認識の認識精度を向上させることができると共に、特定向けにカスタマイズされた言語モデルを、ネットワークを介して受信することにより、ユーザが複数の異なる照合手段を利用する場合でも適切な言語モデルを利用し、高い認識精度を得ることができるという効果がある。

【０１８１】この発明によれば、照合手段が音声認識の際に参照する単語を登録したユーザ辞書を備え、言語モデル更新手段が、ネットワークを介して接続された言語モデル管理サーバから、更新された言語データにより構築された、ユーザ辞書に依存した言語モデルを受信し、照合手段が音声認識の際に参照する言語モデルを、受信した言語モデルにより更新することにより、ユーザ辞書が大きくなった場合でも、ユーザに大きな負担をかけることなく、音声認識の認識精度を向上させることができるという効果がある。

【０１８２】この発明によれば、言語モデル更新手段が、ネットワークを介して接続された言語モデル管理サーバから、更新された言語データにより構築された、音声認識を行うユーザが利用したテキストに依存した言語モデルを受信し、照合手段が音声認識の際に参照する言語モデルを、受信した言語モデルにより更新することにより、ユーザに大きな負担をかけることなく、音声認識の認識精度を向上させることができるという効果がある。

【０１８３】この発明によれば、音声の音響的な観測値系列の確率を求める音響モデルと、音声信号を入力し上記音響モデルを参照して音声認識を行い、認識結果を出力する照合手段と、音響モデルを特定するＩＤを取得する音響モデルＩＤ取得手段と、取得したＩＤを読み出し、入力された音声信号から適応化用の音声データを取得し、読み出したＩＤ及び取得した適応化用の音声データを、ネットワークを介して接続された音響モデル管理サーバに送信する適応化用音声取得手段と、音響モデル管理サーバから、ＩＤに対応する適応化用の音声データにより適応化された適応化済み音響モデルを受信し、照合手段が音声認識の際に参照する音響モデルを、受信した適応化済み音響モデルにより更新する音響モデル更新手段とを備えたことにより、ユーザに大きな負担をかけることなく、音声認識の認識精度を向上させることができると共に、ユーザに適応化した適応化済み音響モデルにより、ネットワークを介して音響モデルを更新することで、ユーザが複数の異なる照合手段を利用する場合でも、適切な音響モデルを利用し、高い認識精度を得ることができるという効果がある。

【０１８４】この発明によれば、更新された音響データを取得する音響データ取得手段と、外部からの音響モデル更新指令を受けて更新された音響データを読み出し、音声の音響的な観測値系列の確率を求める音響モデルを構築する音響モデル構築手段と、音響モデル構築手段により構築された音響モデルを、ネットワークを介して音声認識を行う音声認識装置に送信する音響モデル送信手段とを備えたことにより、ユーザに大きな負担をかけることなく、音声認識の認識精度を向上させることができるという効果がある。

【０１８５】この発明によれば、更新された音響データを取得する音響データ取得手段と、外部からの音響モデル更新指令を受けて、音声認識の際に参照する音響モデルを特定するＩＤを取得する更新音響モデルＩＤ取得手段と、取得したＩＤで指示される特定条件に対応して、更新された音響データを読み出す特定向け音響データ読み出し手段と、読み出した更新された音響データを参照し、特定条件に依存した音響モデルを構築する特定向け音響モデル構築手段と、構築した音響モデルを、ネットワークを介して音声認識装置に送信する音響モデル送信手段とを備えたことにより、ユーザに大きな負担をかけることなく、音声認識の認識精度を向上させることができると共に、特定向けにカスタマイズされた音響モデルを、ネットワークを介して送信することにより、ユーザが複数の異なる音声認識装置を利用する場合でも適切な

音響モデルを利用し、高い認識精度を得ることができるという効果がある。

【０１８６】この発明によれば、音声の音響的な観測値系列の確率を求める、適応化前の初期音響モデルと、ネットワークを介して接続された音声認識装置から送信された、適応化用の音声データと、音声認識装置が音声認識の際に参照する音響モデルを特定するＩＤを受信し、適応化用の音声データを用いて初期音響モデルを適応化し、適応化済み音響モデルを、受信したＩＤに対応付けて適応化済み音響モデル格納手段に格納する音響モデル適応化手段と、外部からの音響モデル更新指令を受けて、ネットワークを介して音声認識装置からＩＤを受信し、受信したＩＤに対応する適応化済み音響モデルを、適応化済み音響モデル格納手段から選択して読み出す適応化済み音響モデル選択手段と、読み出した適応化済み音響モデルを、ネットワークを介して音声認識装置に送信する音響モデル送信手段とを備えたことにより、ユーザに大きな負担をかけることなく、音声認識の認識精度を向上させることができると共に、ユーザに適応化した適応化済み音響モデルにより、ネットワークを介して音響モデルを更新することで、ユーザが複数の異なる音声認識装置を利用する場合でも、適切な音響モデルを利用し、高い認識精度を得ることができるという効果がある。

【０１８７】この発明によれば、更新された言語データを取得する言語データ取得手段と、外部からの言語モデル更新指令を受けて更新された言語データを読み出し、単語列の出現確率を求める言語モデルを構築する言語モデル構築手段と、言語モデル構築手段が構築した言語モデルを、ネットワークを介して音声認識を行う音声認識装置に送信する言語モデル送信手段とを備えたことにより、ユーザに大きな負担をかけることなく、音声認識の認識精度を向上させることができるという効果がある。

【０１８８】この発明によれば、更新された言語データを取得する言語データ取得手段と、外部からの言語モデル更新指令を受けて、音声認識の際に参照する言語モデルを特定するＩＤを取得する更新言語モデルＩＤ取得手段と、ＩＤで指示される特定条件に対応して、更新された言語データを読み出す特定向け言語データ読み出し手段と、読み出した更新された言語データを参照し、特定条件に依存した言語モデルを構築する特定向け言語モデル構築手段と、構築した言語モデルを、ネットワークを介して音声認識装置に送信する言語モデル送信手段とを備えたことにより、ユーザに大きな負担をかけることなく、音声認識の認識精度を向上させることができると共に、特定向けにカスタマイズされた言語モデルを、ネットワークを介して送信することにより、ユーザが複数の異なる音声認識装置を利用する場合でも適切な言語モデルを利用し、高い認識精度を得ることができるという効果がある。

【０１８９】この発明によれば、更新された言語データを取得する言語データ取得手段と、外部からの言語モデル更新指令を受けて、ネットワークを介して接続された音声認識装置が音声認識の際に参照するユーザ辞書を読み出すユーザ辞書読み出し手段と、更新された言語データを読み出し、ユーザ辞書に依存した言語モデルを構築するユーザ辞書依存言語モデル構築手段と、構築した言語モデルを、ネットワークを介して音声認識装置に送信する言語モデル送信手段とを備えたことにより、ユーザ辞書が大きくなった場合でも、ユーザに大きな負担をかけることなく、音声認識の認識精度を向上させることができるという効果がある。

【０１９０】この発明によれば、更新された言語データを取得する言語データ取得手段と、外部からの言語モデル更新指令を受けて、ユーザが利用したテキストを取得するユーザ利用テキスト取得手段と、更新された言語データを読み出し、取得したテキストに依存した言語モデルを構築するユーザ利用テキスト依存言語モデル構築手段と、構築した言語モデルを、ネットワークを介して音声認識装置に送信する言語モデル送信手段とを備えたことにより、ユーザに大きな負担をかけることなく、音声認識の認識精度を向上させることができるという効果がある。

【０１９１】この発明によれば、更新された音響データを取得する第１のステップと、音響モデル更新指令を受けて更新された音響データを読み出し、音響モデルを構築する第２のステップと、構築した音響モデルを、ネットワークを介して送信する第３のステップと、音響モデルを受信し、音声認識の際に参照する音響モデルを、受信した音響モデルにより更新する第４のステップとを備えたことにより、ユーザに大きな負担をかけることなく、音声認識の認識精度を向上させることができるという効果がある。

【０１９２】この発明によれば、更新された言語データを取得する第１のステップと、言語モデル更新指令を受けて更新された言語データを読み出し、言語モデルを構築する第２のステップと、構築した言語モデルを、ネットワークを介して送信する第３のステップと、言語モデルを受信し、音声認識の際に参照する言語モデルを、受信した言語モデルにより更新する第４のステップとを備えたことにより、ユーザに大きな負担をかけることなく、音声認識の認識精度を向上させることができるという効果がある。

【０１９３】この発明によれば、更新された音響データを取得する第１のステップと、音響モデル更新指令を受けて、音声認識の際に参照する音響モデルを特定するＩＤを取得する第２のステップと、ＩＤで指示される特定条件に対応して、更新された音響データを読み出す第３のステップと、更新された音響データを参照し、特定条件に依存した音響モデルを構築する第４のステップと、

構築した音響モデルを、ネットワークを介して送信する第５のステップと、音響モデルを受信し、音声認識の際に参照する音響モデルを、受信した音響モデルにより更新する第６のステップとを備えたことにより、ユーザに大きな負担をかけることなく、音声認識の認識精度を向上させることができると共に、特定向けにカスタマイズされた音響モデルを、ネットワークを介して受信することにより、ユーザが複数の異なる音声認識方法を利用する場合でも適切な音響モデルを利用し、高い認識精度を得ることができるという効果がある。

【０１９４】この発明によれば、更新された言語データを取得する第１のステップと、言語モデル更新指令を受けて、音声認識の際に参照する言語モデルを特定するＩＤを取得する第２のステップと、取得したＩＤで指示される特定条件に対応して、更新された言語データを読み出す第３のステップと、更新された言語データを参照し、特定条件に依存した言語モデルを構築する第４のステップと、構築した言語モデルを、ネットワークを介して送信する第５のステップと、言語モデルを受信し、音声認識の際に参照する言語モデルを、受信した言語モデルにより更新する第６のステップとを備えたことにより、ユーザに大きな負担をかけることなく、音声認識の認識精度を向上させることができると共に、特定向けにカスタマイズされた言語モデルを、ネットワークを介して受信することにより、ユーザが複数の異なる音声認識方法を利用する場合でも適切な言語モデルを利用し、高い認識精度を得ることができるという効果がある。

【０１９５】この発明によれば、更新された言語データを取得する第１のステップと、言語モデル更新指令を受けて、音声認識の際に参照するユーザ辞書を読み出す第２のステップと、更新された言語データを読み出し、ユーザ辞書に依存した言語モデルを構築する第３のステップと、構築した言語モデルを、ネットワークを介して送信する第４のステップと、言語モデルを受信し、音声認識の際に参照する言語モデルを、受信した言語モデルにより更新する第５のステップとを備えたことにより、ユーザ辞書が大きくなった場合でも、ユーザに大きな負担をかけることなく、音声認識の認識精度を向上させることができるという効果がある。

【０１９６】この発明によれば、更新された言語データを取得する第１のステップと、言語モデル更新指令を受けて、ユーザが利用したテキストを取得する第２のステップと、更新された言語データを読み出し、テキストに依存した言語モデルを構築する第３のステップと、構築した言語モデルを、ネットワークを介して送信する第４のステップと、言語モデルを受信し、音声認識の際に参照する言語モデルを、受信した言語モデルにより更新する第５のステップとを備えたことにより、ユーザに大きな負担をかけることなく、音声認識の認識精度を向上させることができるという効果がある。

【０１９７】この発明によれば、音響モデルを特定するＩＤを取得する第１のステップと、取得したＩＤを読み出し、入力された音声信号から適応化用の音声データを取得し、ネットワークを介して、読み出したＩＤ及び取得した適応化用の音声データを送信する第２のステップと、送信した適応化用の音声データを用いて、適応化前の初期音響モデルを適応化し、適応化済み音響モデルを、送信したＩＤに対応付けて格納する第３のステップと、音響モデル更新指令を受けて、ネットワークを介して第１のステップで取得したＩＤを受信し、受信したＩＤに対応する適応化済み音響モデルを、第３のステップで格納している適応化済み音響モデルの中から選択して読み出す第４のステップと、第４のステップで読み出した適応化済み音響モデルを、ネットワークを介して送信する第５のステップと、送信した適応化済み音響モデルを受信し、音声認識の際に参照する音響モデルを、受信した適応化済み音響モデルにより更新する第６のステップとを備えたことにより、ユーザに大きな負担をかけることなく、音声認識の認識精度を向上させることができると共に、ユーザに適応化した適応化済み音響モデルにより、ネットワークを介して音響モデルを更新することで、ユーザが複数の異なる音声認識方法を利用する場合でも、適切な音響モデルを利用し、高い認識精度を得ることができるという効果がある。

【０１９８】この発明によれば、音声認識プログラムを記録した記録媒体で、更新された音響データを取得する音響データ取得機能と、音響モデル更新指令を受けて更新された音響データを読み出し、音響モデルを構築する音響モデル構築機能と、構築した音響モデルを、ネットワークを介して送信する音響モデル送信機能と、音響モデルを受信し、照合機能が音声認識の際に参照する音響モデルを、受信した音響モデルにより更新する音響モデル更新機能とを実現させることにより、ユーザに大きな負担をかけることなく、音声認識の認識精度を向上させることができるという効果がある。

【０１９９】この発明によれば、音声認識プログラムを記録した記録媒体で、更新された言語データを取得する言語データ取得機能と、言語モデル更新指令を受けて、上記言語データ取得機能が取得した更新された言語データを読み出し、言語モデルを構築する言語モデル構築機能と、上記言語モデル構築機能が構築した言語モデルを、ネットワークを介して送信する言語モデル送信機能と、上記言語モデル送信機能が送信した言語モデルを受信し、上記照合機能が音声認識の際に参照する言語モデルを、受信した言語モデルにより更新する言語モデル更新機能とを実現させることにより、ユーザに大きな負担をかけることなく、音声認識の認識精度を向上させることができるという効果がある。

【０２００】この発明によれば、音声認識プログラムを記録した記録媒体で、更新された音響データを取得する

音響データ取得機能と、音響モデル更新指令を受けて、音響モデルを特定するＩＤを取得する更新音響モデルＩＤ取得機能と、取得したＩＤで指示される特定条件に対応して、更新された音響データを読み出す特定向け音響データ読み出し機能と、更新された音響データを参照し、特定条件に依存した音響モデルを構築する特定向け音響モデル構築機能と、音響モデルを、ネットワークを介して送信する音響モデル送信機能と、音響モデルを受信し、照合機能が音声認識の際に参照する音響モデルを、受信した音響モデルにより更新する音響モデル更新機能とを実現させることにより、ユーザに大きな負担をかけることなく、音声認識の認識精度を向上させることができると共に、特定向けにカスタマイズされた音響モデルを、ネットワークを介して受信することにより、ユーザが複数の照合機能を利用する場合でも適切な音響モデルを利用し、高い認識精度を得ることができるという効果がある。

【０２０１】この発明によれば、音声認識プログラムを記録した記録媒体で、更新された言語データを取得する言語データ取得機能と、言語モデル更新指令を受けて、言語モデルを特定するＩＤを取得する更新言語モデルＩＤ取得機能と、取得したＩＤで指示される特定条件に対応して、更新された言語データを読み出す特定向け言語データ読み出し機能と、更新された言語データを参照し、特定条件に依存した言語モデルを構築する特定向け言語モデル構築機能と、構築した言語モデルを、ネットワークを介して送信する言語モデル送信機能と、言語モデルを受信し、照合機能が音声認識の際に参照する言語モデルを、受信した言語モデルにより更新する言語モデル更新機能とを実現させることにより、ユーザに大きな負担をかけることなく、音声認識の認識精度を向上させることができると共に、特定向けにカスタマイズされた言語モデルを、ネットワークを介して受信することにより、ユーザが複数の照合機能を利用する場合でも適切な言語モデルを利用し、高い認識精度を得ることができるという効果がある。

【０２０２】この発明によれば、音声認識プログラムを記録した記録媒体で、更新された言語データを取得する言語データ取得機能と、言語モデル更新指令を受けて、ユーザ辞書を読み出すユーザ辞書読み出し機能と、更新された言語データを読み出し、ユーザ辞書に依存した言語モデルを構築するユーザ辞書依存言語モデル構築機能と、構築した言語モデルを、ネットワークを介して送信する言語モデル送信機能と、言語モデルを受信し、照合機能が音声認識の際に参照する言語モデルを、受信した言語モデルにより更新する言語モデル更新機能とを実現させることにより、ユーザ辞書が大きくなった場合でも、ユーザに大きな負担をかけることなく、音声認識の認識精度を向上させることができるという効果がある。

【０２０３】この発明によれば、音声認識プログラムを記録した記録媒体で、更新された言語データを取得する言語データ取得機能と、言語モデル更新指令を受けて、ユーザが利用したテキストを取得するユーザ利用テキスト取得機能と、更新された言語データを読み出し、テキストに依存した言語モデルを構築するユーザ利用テキスト依存言語モデル構築機能と、構築した言語モデルを、ネットワークを介して送信する言語モデル送信機能と、言語モデルを受信し、照合機能が音声認識の際に参照する言語モデルを、受信した言語モデルにより更新する言語モデル更新機能とを実現させることにより、ユーザに大きな負担をかけることなく、音声認識の認識精度を向上させることができるという効果がある。

【０２０４】この発明によれば、音声認識プログラムを記録した記録媒体で、音響モデルを特定するＩＤを取得する音響モデルＩＤ取得機能と、取得したＩＤを読み出し、入力された音声信号から適応化用の音声データを取得し、ネットワークを介して、読み出したＩＤ及び取得した適応化用の音声データを送信する適応化用音声取得機能と、送信した適応化用の音声データを用いて、適応化前の初期音響モデルを適応化し、適応化済み音響モデルを、送信したＩＤに対応付けて格納する音響モデル適応化機能と、音響モデル更新指令を受けて、ネットワークを介して音響モデルＩＤ取得機能が取得したＩＤを受信し、受信したＩＤに対応する適応化済み音響モデルを、音響モデル適応化機能が格納した適応化済み音響モデルの中から選択して読み出す適応化済み音響モデル選択機能と、読み出した適応化済み音響モデルを、ネットワークを介して送信する音響モデル送信機能と、適応化済み音響モデルを受信し、照合機能が音声認識の際に参照する音響モデルを、受信した適応化済み音響モデルにより更新する音響モデル更新機能とを実現させることにより、ユーザに大きな負担をかけることなく、音声認識の認識精度を向上させることができると共に、ユーザに適応化した適応化済み音響モデルにより、ネットワークを介して音響モデルを更新することで、ユーザが複数の異なる照合機能を利用する場合でも、適切な音響モデルを利用し、高い認識精度を得ることができるという効果がある。

【図面の簡単な説明】

【図１】　この発明の実施の形態１による音声認識装置の構成を示すブロック図である。

【図２】　この発明の実施の形態１による音響モデルの更新処理の手順を示すフローチャートである。

【図３】　この発明の実施の形態１による言語モデルの更新処理の手順を示すフローチャートである。

【図４】　この発明の実施の形態２による音声認識装置の構成を示すブロック図である。

【図５】　この発明の実施の形態２による言語モデルの更新処理の手順を示すフローチャートである。

【図６】　この発明の実施の形態３による音声認識装置

の構成を示すブロック図である。

【図７】　この発明の実施の形態３による言語モデルの更新処理の手順を示すフローチャートである。

【図８】　この発明の実施の形態４による音声認識装置の構成を示すブロック図である。

【図９】　この発明の実施の形態４による言語モデルの更新処理の手順を示すフローチャートである。

【図１０】　この発明の実施の形態５による音声認識装置の構成を示すブロック図である。

【図１１】　この発明の実施の形態５による音響モデルの更新処理の手順を示すフローチャートである。

【図１２】　従来の音声認識装置の構成を示すブロック図である。

【図１３】　従来の音声認識処理の手順を示すフローチャートである。

【図１４】　従来の音声認識装置の構成を示すブロック図である。

【図１５】　従来の音声認識装置の構成を示すブロック図である。

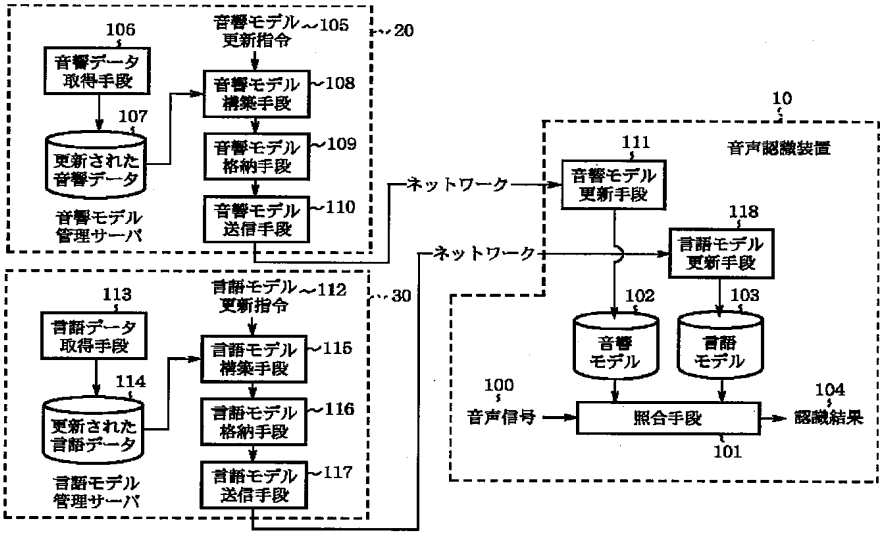
【図１６】　ユーザ辞書６０１の構成例を示す図である。

【図１７】　従来の音声認識処理の手順を示すフローチャートである。

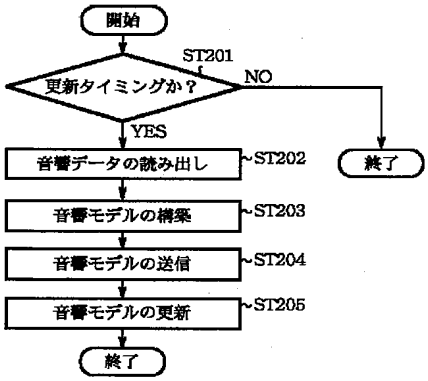
＊【符号の説明】

１０　音声認識装置、２０　音響モデル管理サーバ、３０　言語モデル管理サーバ、１００　音声信号、１０１　照合手段、１０２　音響モデル、１０３　言語モデル、１０４　認識結果、１０５　音響モデル更新指令、１０６　音響データ取得手段、１０７　更新された音響データ、１０８　音響モデル構築手段、１０９　音響モデル格納手段、１１０　音響モデル送信手段、１１１　音響モデル更新手段、１１２　言語モデル更新指令、１１３　言語データ取得手段、１１４更新された言語データ、１１５　言語モデル構築手段、１１６　言語モデル格納手段、１１７　言語モデル送信手段、１１８　言語モデル更新手段、４０１更新言語モデルＩＤ取得手段、４０２　特定向け言語データ読み出し手段、４０３　特定向け言語モデル構築手段、６０１　ユーザ辞書、６０２　ユーザ辞書読み出し手段、６０３　ユーザ辞書依存言語モデル構築手段、８０１　ユーザ利用テキスト取得手段、８０２　ユーザ利用テキスト格納手段、８０３　ユーザ利用テキスト依存言語モデル構築手段、１００１　音響モデルＩＤ取得手段、１００２　適応化用音声取得手段、１００３　初期音響モデル、１００４　音響モデル適応化手段、１００５　適応化済み音響モデル格納手段、１００６　適応化済み音響モデル選択手段。

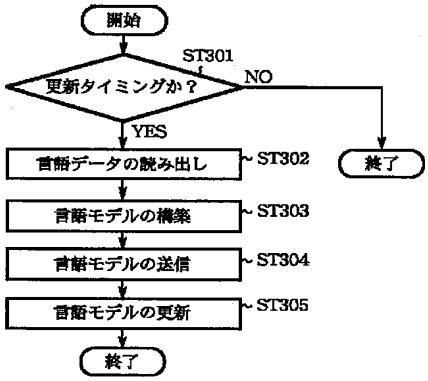
【図１】



【図２】



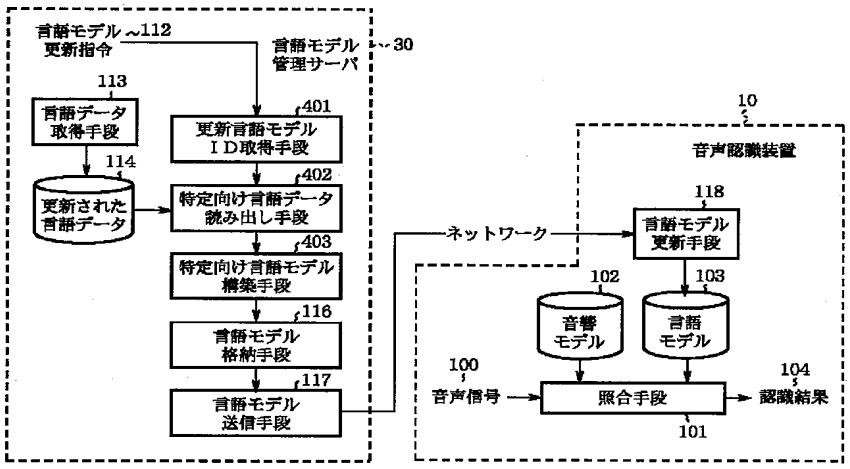
【図３】



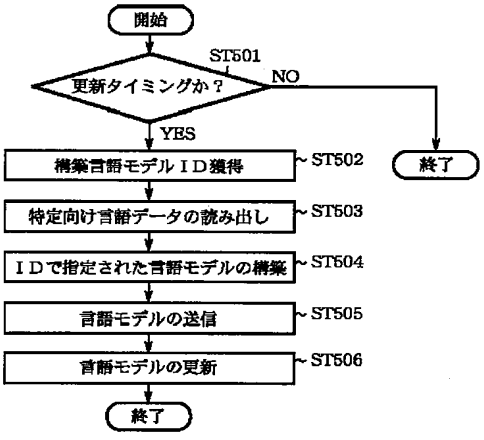
【図１６】

漢字表記	よみ
音素	オンソ
音声認識	オンセイニンシキ
多変量解析	タヘンリョウカイセキ
チューリング	チューリング
⋮	⋮

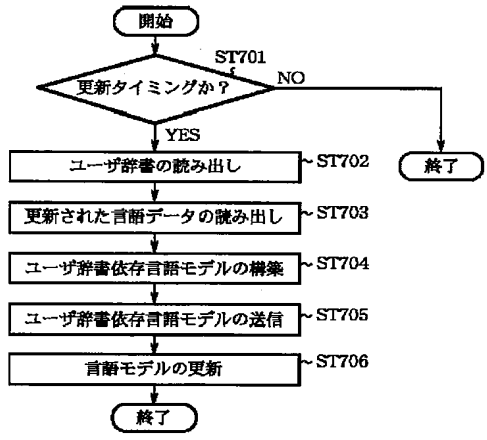
【図４】



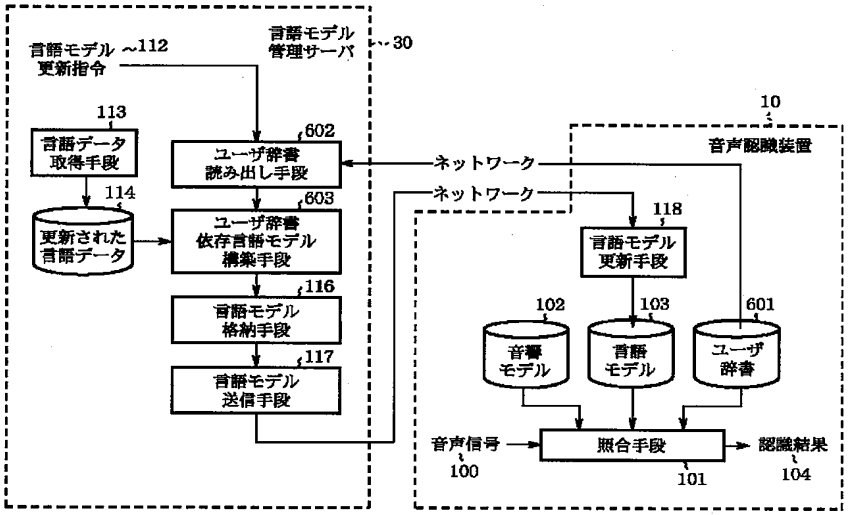
【図５】



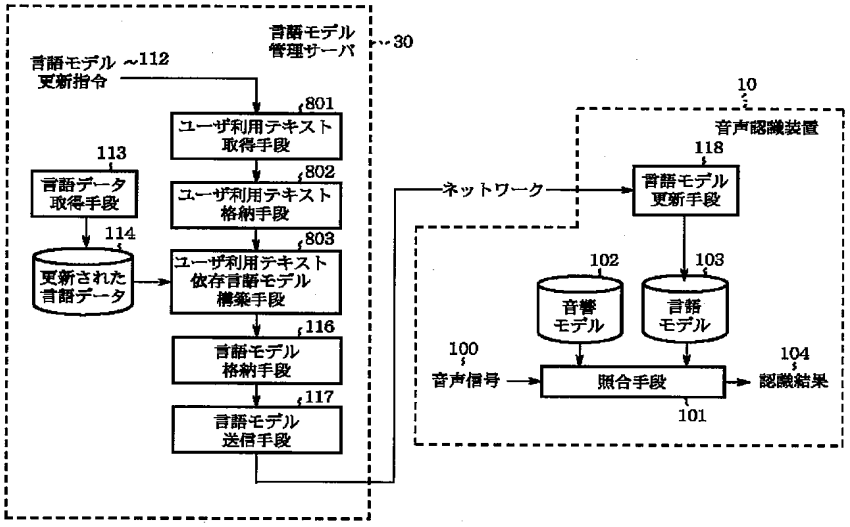
【図７】



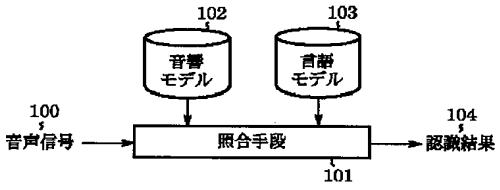
【図６】


言語モデル
更新指令
112
言語モデル
管理サーバ
30
10
音声認識装置
113
言語データ
取得手段
602
ユーザ辞書
読み出し手段
ネットワーク
ネットワーク
114
更新された
言語データ
603
ユーザ辞書
依存言語モデル
構築手段
118
言語モデル
更新手段
116
言語モデル
格納手段
102
音響
モデル
103
言語
モデル
601
ユーザ
辞書
117
言語モデル
送信手段
音声信号
100
照合手段
101
認識結果
104

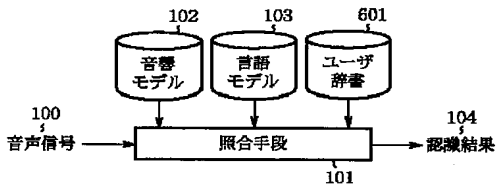
【図８】


言語モデル
更新指令
112
言語モデル
管理サーバ
30
10
音声認識装置
801
ユーザ利用テキスト
取得手段
113
言語データ
取得手段
802
ユーザ利用テキスト
格納手段
ネットワーク
118
言語モデル
更新手段
114
更新された
言語データ
803
ユーザ利用テキスト
依存言語モデル
構築手段
102
音響
モデル
103
言語
モデル
116
言語モデル
格納手段
100
音声信号
照合手段
104
認識結果
101
117
言語モデル
送信手段

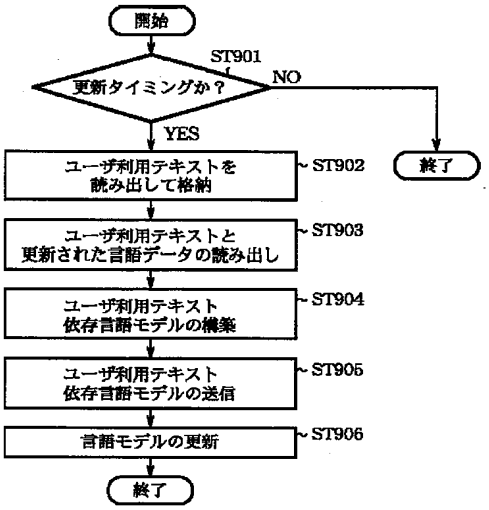
【図１２】


102
103
音響
モデル
言語
モデル
100
音声信号
照合手段
104
認識結果
101

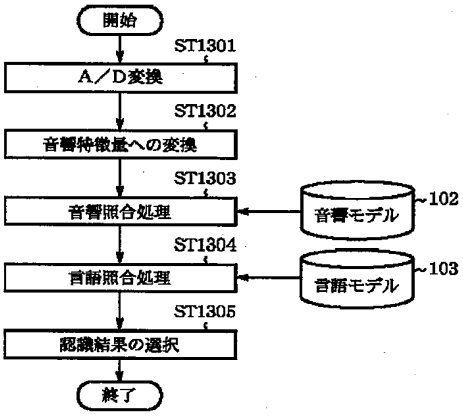
【図１５】


102
103
601
音響
モデル
言語
モデル
ユーザ
辞書
100
音声信号
照合手段
104
認識結果
101

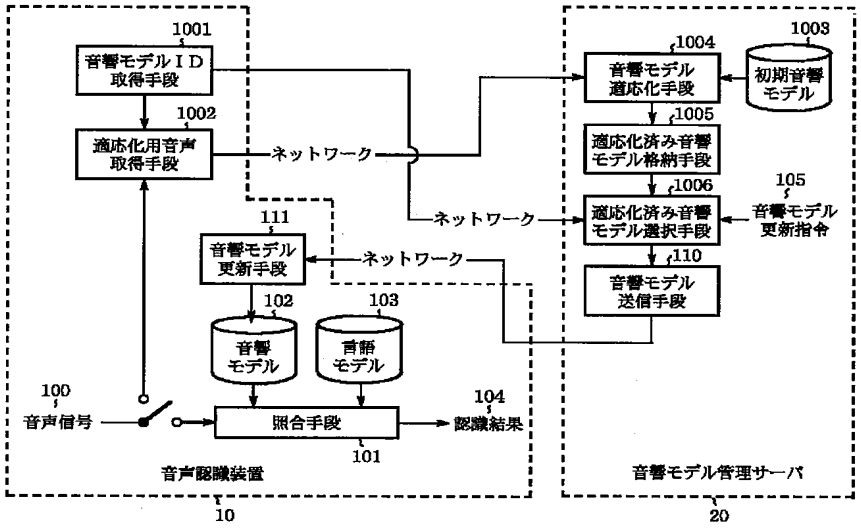
【図９】


開始
ST901
更新タイミングか？
NO
YES
終了
ユーザ利用テキストを読み出して格納
ST902
ユーザ利用テキストと更新された言語データの読み出し
ST903
ユーザ利用テキスト依存言語モデルの構築
ST904
ユーザ利用テキスト依存言語モデルの送信
ST905
言語モデルの更新
ST906
終了

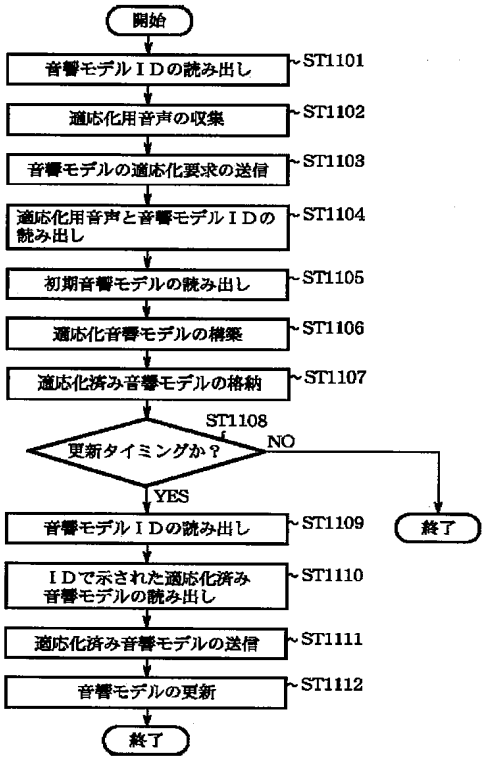
【図１３】


開始
ST1301
A／D変換
ST1302
音響特徴量への変換
ST1303
音響照合処理
音響モデル
102
ST1304
言語照合処理
言語モデル
103
ST1305
認識結果の選択
終了

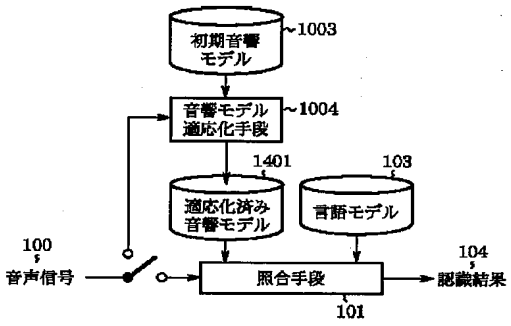
【図１０】


1001
音響モデルＩＤ取得手段
1002
適応化用音声取得手段
ネットワーク
111
音響モデル更新手段
ネットワーク
102
音響モデル
103
言語モデル
100
音声信号
照合手段
101
104
認識結果
音声認識装置
10
1003
初期音響モデル
1004
音響モデル適応化手段
1005
適応化済み音響モデル格納手段
1006
適応化済み音響モデル選択手段
105
音響モデル更新指令
110
音響モデル送信手段
ネットワーク
音響モデル管理サーバ
20

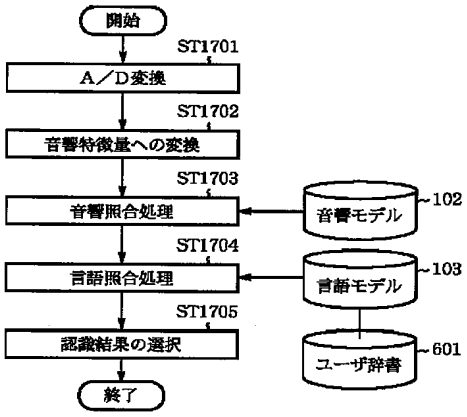
【図１１】


開始
音響モデルＩＤの読み出し
ST1101
適応化用音声の収集
ST1102
音響モデルの適応化要求の送信
ST1103
適応化用音声と音響モデルＩＤの読み出し
ST1104
初期音響モデルの読み出し
ST1105
適応化音響モデルの構築
ST1106
適応化済み音響モデルの格納
ST1107
ST1108
更新タイミングか？
NO
YES
終了
音響モデルＩＤの読み出し
ST1109
ＩＤで示された適応化済み音響モデルの読み出し
ST1110
適応化済み音響モデルの送信
ST1111
音響モデルの更新
ST1112
終了

【図１４】


初期音響モデル
1003
音響モデル適応化手段
1004
1401
103
適応化済み音響モデル
言語モデル
100
音声信号
照合手段
104
認識結果
101

【図１７】


開始
ST1701
A／D変換
ST1702
音響特徴量への変換
ST1703
音響照合処理
音響モデル
102
ST1704
言語照合処理
言語モデル
103
ST1705
認識結果の選択
ユーザ辞書
601
終了